The Clustering of Galaxies in the SWIRE Survey

Mark Frost

Astronomy Centre, University of Sussex

Submitted for the degree of

Doctor of Philosophy (D.Phil.)

September 2009

Supervision by:

Dr. Seb Oliver

Declaration

I have read and understood the definition of 'Plagiarism' as set out in the General Assessment Handbook for Masters and Postgraduate Diploma and Postgraduate Certificate Candidates under item 12.

The work related to the IRAC and MIPS masking to constrain selection effects, outlined in Chapter 3, was published in Monthly Notices for the Royal Astronomical Society:

Specific star-formation and the relation to stellar mass from $0 < z < 2$ *as seen in the far-infrared at 70 and 160*μm, Oliver et. al. 2009, (submitted)

The work outlined in Chapter 6 was published in the Astrophysical Letters:

A Pilot Search for Population III Supernova Candidates in the Spitzer/IRAC dark field, Frost et. al. 2009, Astrophysical Journal Letters 698 (2009) L68-L71.

Signed..... Date.....

Mark Frost

Abstract

Despite the coherent cosmological framework provided by the λCDM model that astronomers have to work within, there are still a lot of unanswered questions regarding galaxy formation and evolution. Measuring the clustering of galaxies can provide information about the different environments that different types of galaxies reside in. Also, measuring the clustering of similar samples of galaxies at different redshifts can provide insights into how galaxies have evolved over time.

Previous clustering analyses, particularly at high redshift, have often been restricted to galaxy samples which are small, selected on observable properties and/or contain an unknown mixture of different spectral types. Small samples lead to limited statistics and the inability to break the sample into interesting subsamples based on properties, e.g. by luminosity or star formation rate. Selecting samples based on observable properties leads to varying intrinsic properties with redshift and hence makes interpreting the evolution of clustering difficult. Mixing spectral types makes it impossible to separate the contribution to the clustering signal from early and late-type galaxies which tend to cluster very differently.

This thesis overcomes some of the limitations of earlier clustering analyses by using the Spitzer Wide area InfraRed Extragalactic (SWIRE) photometric redshift catalogue of Rowan-Robinson et al.(2008) to measure the clustering of galaxies. The SWIRE catalogues covered multiple fields and large volumes providing large samples of galaxies over $0.1 < z < 1.5$. The template fitting procedure also provides spectral classifications as well as intrinsic properties such as stellar mass and star formation rate estimates. The clustering of elliptical and spiral galaxies detected in SWIRE is measured as a function of stellar mass over $0.1 \leq z \leq 1.5$. The clustering of spiral galaxies selected on star formation rate is also investigated over the same redshift range. Such measurements can help constrain theories of galaxy evolution.

Another Spitzer dataset, the dark field, is used in an attempt to place one of the first observational constraints on the detection rate of population III supernovae. The dark field is an extragalactic data set with repeat imaging on a monthly basis over a baseline of approximately 2 years. The unprecedented depth and multi-epochal nature of this data makes it ideal for a first foray into trying to detect supernovae from the first stars.

Acknowledgements

First of all I would like to thank everyone in the Astronomy Centre. I don't think there is a friendlier, more supportive department in the country and I am grateful for my time there. I would like to thank my supervisor Seb Oliver for always being positive and enthusiastic, an attitude which served as great motivation on countless occasions and helped me through the challenging times when the selection effects seemed too much. I would also like to thank the rest of Seb's research group, past and present, for their help and support over the last few years. Your insights and creativity have helped to shape the research within these pages. My friends at Sussex have made this journey fun and I know I have forged numerous friendships that will stand the test of time.

Finally I must thank my family, your support has been unwavering and I am so grateful for everything you have done for me. Last but not least, Eryn - thank you for your patience as I spent countless evenings 'finishing one last calculation' and for listening intently to my ramblings about dark matter. None of it went unnoticed.

Official acknowledgements:-

- A pair-counting code utilizing kd-tree methods, the NPT code, was provided by The Auton Lab at Carnegie Mellon University's School of Computer Science (http://www.autonlab.org) and we thank Andrew Moore et al. for making this software public. It's use is adopted throughout this thesis.

- This thesis makes use of observations made with the Spitzer Space Telescope, operated by the Jet Propulsion Laboratory, California Institute of Technology under NASA contract 1407.

- This thesis makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

- Funding to do the research in this thesis was provided by STFC studentship PPA/S/S2005/04270.

Dedication

To my Parents:

Without your encouragement and support I would never have had the opportunity to study astronomy. I cannot thank you enough.

All my love to both of you

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Universe as we know it

The scientific disciplines of astronomy and cosmology have long been studied by man although not always under those names. However it has only been in the last 80 years that a coherent model, the Big Bang model, for the creation and evolution of the Universe has emerged. There are three key pieces of observational evidence, what are sometimes referred to as the three pillars, which the Big Bang theory now adequately explains.

The oldest piece is the recessional velocities of galaxies. In 1923/1924 (although not published until 1929) Edwin Hubble discovered that all galaxies are moving away from each other and that the recessional velocity, V, and the distance, D, to a galaxy are directly proportional to each other.

$$V = H(t)D \tag{1.1}$$

The proportionality constant $H(t)$ is known as Hubble's parameter. This showed that the Universe is expanding and hence implied that it must have been smaller in the past. Extrapolating backwards in time we can infer the Universe started as a singularity at time zero. In the 1920's Alexander Friedmann developed a dynamic equation which describes the expansion of the Universe.

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} \tag{1.2}$$

where $a(t)$ is the *scale factor* which measures the expansion rate, ρ is the density of the Universe and k represents the curvature of the Universe which can take three forms; spher-

ical, flat or hyperbolic corresponding to k being positive, zero or negative, respectively. The Hubble parameter can then be expressed as,

$$H(t) = \frac{\dot{a}(t)}{a(t)} \tag{1.3}$$

The second main piece of evidence is the Cosmic Microwave Background (CMB) which was detected for the first time in 1964 by Penzias and Wilson. The Big Bang theory says the early Universe was a hot dense plasma of particles where photons were created and destroyed. During this period the Universe was in thermal equilibrium giving the photons a blackbody spectrum. As the Universe cooled, photons continued to be Thomson scattered off free electrons thereby making the Universe opaque. Approximately 379,000 years after the Big Bang the Universe had cooled to a point where atoms were able to form capturing the free electrons. Photons were then able to propagate freely through the Universe, initially with a temperature of $\sim$ 3000K. Due to the expansion of the Universe the photons energy decreased over time leading to what we see today as the CMB. It's almost uniform temperature and shape as a blackbody accurately matched predictions made by Alpher and Herman in 1948 of the hot initial state of the early Universe resulting from a Big Bang.

The third main piece of observational evidence is the measured abundances of light elements. It was originally assumed that all elements except hydrogen were created in stars but consideration of the nuclear fusion mechanisms cannot account for the measured abundances of light elements such as deuterium, helium-3, helium-4 and lithium. However, these abundances can be accounted for quite well in the Big Bang theory. When the Universe was 3 minutes old it's temperature had fallen to $\sim 10^9$ K allowing light elements to form through the process of nucleosynthesis. As the Universe continued to expand, the density and temperature continued to fall and when the Universe was approximately 20 minutes old both had fallen to such a level that fusion was no longer possible. This prevented elements heavier than beryllium forming and hence allowed the lighter elements to survive. The predicted abundances of light elements depends on the ratio of baryons to photons which has been accurately measured from the CMB by WMAP (Cyburt et al.(2003)). Using this ratio, the calculated abundances of helium-4, helium-3 and deuterium are found to be in good agreement with observed abundances.

Despite explaining these three key observations the Big Bang model requires some additional theories to fully explain the observed Universe. The density of the Universe

today is very close to it's critical value and hence the Universe is close to being flat. Therefore the Universe must have been even closer to flat in the past as any small deviation from $k = 0$ quickly leads to a Universe with curvature. Such a set of circumstances is perhaps too convenient and this became known as the *flatness problem*. The density of the Universe is usually expressed as the density parameter

$$\Omega(t) = \frac{\rho}{\rho_c} \tag{1.4}$$

where ρ_c is the critical density which is the density the Universe would have if it were flat and is defined as

$$\rho_c(t) = \frac{3H^2}{8\pi G} \tag{1.5}$$

The uniform appearance of the CMB on very large scales is not possible to explain as regions of the sky separated by large distances could not have been in causal contact, even in the early Universe. This became known as the *horizon problem*. Both of these problems were resolved by Alan Guth who proposed the theory of *inflation* - an early period of rapid expansion, defined as $\ddot{a} > 0$, which occurred when the Universe was 10^{-37} sec old. This allows the early Universe, prior to inflation, to have been sufficiently small for equilibrium to have been reached and since to have expanded to such a size that our observable Universe is effectively flat. A constant, Λ, can be added into the Friedmann equation to parameterise the effects of inflation,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda}{3} \tag{1.6}$$

It is believed that the Universe is currently undergoing another era of accelerated expansion. The decay rate of the light curves of type Ia supernovae can be used to calibrate the peak luminosity. The progenitors of such supernovae are thought to be carbon-oxygen white dwarfs accreting mass in a binary system. The white dwarf reaches the Chandrasekhar limit and explodes. The peak brightness of type Ia supernovae are seen to become progressively dimmer as we move to higher redshifts. This dimming is attributed to the Universe undergoing another era of accelerated expansion (Riess et al.(1998), Riess et al.(2004), Perlmutter et al.(1999)). The expansion is thought to be due to *dark energy*, the nature of which is unknown but it is thought to permeate the whole universe.

The last, and arguably the most important, component of the Universe is *dark matter*. The first person to infer the existence of dark matter was Fritz Zwicky in 1933. He measured the mass of the Coma cluster by studying the motion of galaxies on its periphery.

His estimate for the mass of dark matter present was 400 times greater than the amount of mass seen in luminous matter. Subsequent evidence includes velocity rotation curves of galaxies, orbital velocities of galaxies in clusters, gravitational lensing by clusters and the temperature distribution of hot gas in galaxies and clusters all of which infer the presence of more mass than what can be seen. It is now commonly accepted that the Universe is permeated by dark matter and it is assumed to only interact with baryonic matter through gravity (Blumenthal et al.(1984)).

The different components of the Universe are often expressed as density parameters with the total energy density of the Universe expressed as

$$\Omega_{tot} = \Omega_b + \Omega_{DM} + \Omega_\Lambda \tag{1.7}$$

where Ω_b and Ω_{DM} are the total baryon and dark matter fractions. Ω_Λ is the cosmological constant which represents the fraction associated with dark energy. From WMAP (Bennett et al.(2003)) measurements of the CMB, combined with constraints from type Ia supernovae and baryon acoustic oscillations, we know the relative contributions to Ω_{tot} are $\Omega_b = 0.0456 \pm 0.0015$, $\Omega_{DM} = 0.228 \pm 0.013$ and $\Omega_\Lambda = 0.726 \pm 0.015$ (Hinshaw et al.(2009)). But how did the Universe evolve from it's (almost) homogeneous early state to contain structures such as galaxies, clusters, superclusters, voids and filaments that we observe today?

1.1.1 The Origin of Large Scale Structure

Examining the positions of galaxies across the sky on large scales ($\sim$ 100's of Mpc) reveals an interesting trait in their distribution. Galaxies are not randomly distributed throughout the Universe but instead trace out a complex 'web-like' structure composed of filaments of galaxies joining clusters or superclusters at each conflux. This structure is often termed the *large scale structure* of the Universe and measuring such structure is often referred to as the *clustering* of galaxies (see Section 1.3). The positioning of galaxies is dictated by their gravitational interaction with dark matter. Hence it is the underlying dark matter density field which plays the defining role in the large scale structure of the Universe. Cosmological models assume these dark matter structures began as Gaussian quantum fluctuations in the dark matter density field in the early Universe. As we consider ever larger scales the amplitude of these fluctuations decreases ensuring that the Universe remains smooth on the largest scales and hence satisfies the cosmological principle - that the Universe is

homogeneous and isotropic. A fluctuation can be characterized by it's density contrast,

$$\delta(\mathbf{x}) \equiv \frac{\delta\rho(\mathbf{x})}{\bar{\rho}} = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \tag{1.8}$$

where $\bar{\rho}$ is the average density of the Universe and $\rho(\mathbf{x})$ is the perturbation density. At early epochs $\rho(\mathbf{x}) \ll 1$ for all scales, which satisfies homogeneity and the perturbations grow in a *linear* fashion as their physical size grows with the expansion of the Universe proportional to $a(t)$. Considering the fluctuations in Fourier space the Fourier expansion of $\delta(\mathbf{x})$ is

$$\delta(\mathbf{x}) = \frac{V}{(2\pi)^3} \int \delta_{\mathbf{k}} e^{-i\mathbf{k}.\mathbf{x}} d^3k \tag{1.9}$$

$$\delta_{\mathbf{k}} = \frac{1}{V} \int \delta(\mathbf{x}) e^{i\mathbf{k}.\mathbf{x}} d^3x \tag{1.10}$$

The Fourier modes δ_k evolve independently whilst in the linear regime hence the density fluctuations are written in terms of the power spectrum

$$P(\mathbf{k}) = |\delta_{\mathbf{k}}|^2 \tag{1.11}$$

It is predicted from inflationary theory that the power spectrum will be a power law.

$$P(\mathbf{k}) \propto k^{-n} \tag{1.12}$$

Over time these fluctuations grew through accretion via gravity and as $\delta \rightarrow 1$, the linear approximation is no longer valid, and the perturbations separate from the expansion and collapse to form a bound selfgravitating spherical structures termed *haloes*. Subsequent mergers of haloes lead to ever larger structures in a process known as *hierarchical structure formation*. Baryonic matter is attracted into the gravitational potential wells of these structures initially forming the first stars as the baryonic matter cools and condenses, e.g Cole et al.(2000), which eventually leads to the formation of galaxies (White & Rees(1978)). The filamentary large scale structure can be explained as a consequence of the gravitational instability of the cold dark matter (CDM) occurring preferentially along the shortest axis of 3D and 2D protostructures (the Zeldovich pancakes).

Despite the strong gravitational interplay between galaxies and dark matter, their distributions are not the same. More strongly clustered galaxies are said to be *biased* (Kaiser(1984)) with respect to the dark matter distribution. The relationship between galaxy and dark matter density fluctuations, usually termed the galaxy bias, b, is given by,

$$b = \frac{\delta_{gal}(x)}{\delta_{DM}(x)} \tag{1.13}$$

Kauffmann et al.(1997).

Dark matter haloes can be accurately modelled by N-body simulations whereas modeling galaxy formation and evolution is much more complex. On large scales ($\gtrsim$ 3 Mpc) theory and simulations have shown (e.g. Jenkins et al.(1998), Weinberg et al.(2004), Springel et al.(2006)) the clustering amplitude of dark matter decreases steadily with redshift. On such scales structure growth is largely driven by gravitational interactions between separate haloes of dark matter (inter-halo clustering). Galaxy clustering would similarly decrease with redshift in proportion to the underlying dark matter if the galaxy distribution was not biased. However, on small scales ($\lesssim$ 3 Mpc) the non-linear evolution that small scales underwent and physical processes present throughout galaxy evolution such as star formation, supernovae, AGN, magnetic fields etc complicate the structure formation process (intra-halo clustering). The actual measured clustering amplitudes are a complicated interplay between the underlying dark matter and the physical processes inherent to galaxy formation and evolution.

The paradigmatic scenario of cosmic structure formation and evolution of the Universe, incorporating the Big Bang model, is commonly known as Λ Cold Dark Matter (ΛCDM) where the Λ represents the accelerating expansion, see Equation 1.6. Despite this coherent cosmological framework there are still a lot of unanswered questions regarding galaxy formation and evolution.

1.2 Galaxies, their Evolution and Environment

Observations of the Universe show that galaxies can be broadly separated into two different types; elliptical and spiral galaxies (e.g. Strateva et al.(2001), Blanton et al.(2003)). It is important to understand the typical characteristics of each type of galaxy and the variation in properties such as age, size, stellar mass content and star formation rates that they display.

1.2.1 Elliptical Galaxies

Elliptical galaxies range in diameter from 0.1 kpc to 100 kpc and in stellar mass from $M_\star \sim 10^7 - 10^{13} M_\odot$. They tend to be comprised of old stellar populations which gives

them their predominantly red colour although there is evidence they contain young stellar populations (Trager et al.(2000)). The light emitted from these galaxies is smoothly distributed throughout the galaxy. They contain very little interstellar medium and hence have very low star formation rates. The degree of ellipticity varies greatly from circular E0 galaxies to highly elongated E7 galaxies.

1.2.2 Spiral Galaxies

The distinguishing features of a spiral galaxy are it's spiral arms and a central nucleus or bulge. The bulge contains mainly old stars and is located at the centre of a rotating disk. The disk can range in size from 5-100 kpc. The spiral arms are comprised of lots of gas and dust and therefore contain regions of on-going star formation producing predominantly young blue stellar populations. Spiral galaxies are usually sub-catagorized into Sa, Sb, Sc, Sd and Sm. a-d indicates the size of the bulge and the tightness of the spiral arms with Sa having the tightest arms and largest bulge and Sd galaxies having almost no bulge. Sm galaxies are *magellanic sprials* which are similar to the Large Magellanic Cloud and usually have one truncated arm. Spiral galaxies can also possess a central bar producing barred spirals denoted SBa, SBb etc. Typical stellar mass contents range from $M_\star \sim 10^9 - 10^{12} M_\odot$.

1.2.3 Starburst Galaxies

Starburst galaxies can be any type of galaxy that have recently experienced an increased rate in the production of stars. This can be confined to specific regions of the galaxy or occur across the whole galaxy. There are are two main stimulii which cause the bursts of star formation. Interactions between neighbouring galaxies can cause the gas within the galaxies to be disturbed. This can lead to collisions between gas clouds, and subsequently star formation. The other main way to begin starbursts is through supernova explosions. These disturb and heat the surrounding gas, leading to star formation. As a result of these processes starburst galaxies tend to be very luminous and have a disturbed appearance.

1.2.4 Galaxy Environment and Evolution

Understanding how the different types of galaxies, with their large variation in properties, have formed and evolved is a complex problem. Large extra-galactic surveys, which have

become common place over the past decade, have provided a plethora of data which can be used to further our understanding of the environments galaxies reside in and how they have evolved. Some important observations relating to these aspects of extra-galactic astronomy are highlighted below.

Different types of galaxies reside in different environments. Elliptical galaxies are preferentially found in the densest regions of the Universe such as groups and clusters whereas spiral galaxies constitute the bulk of field galaxies in sparser regions. This is more commonly known as the morphology-density or colour-density relation which was first quantified by Dressler(1980). This bi-modality was later detected by Blanton et al.(2003) and Madgwick et al.(2002) at $z < 0.3$. Coil et al.(2006) later showed that the colour-density relation was already in place by $z \sim 1$. Cucciati et al.(2006) (VVDS) and Cooper et al.(2006) (DEEP2) similarly found the relation to be in place at $z = 1$ but also for it to have disappeared at $z \sim 1.3$. However, the colour-density relation was detected at $z \sim 2.5$ by Daddi et al.(2003), Grazian et al.(2006) and Quadri et al.(2007). It is highly unlikely the colour-density relation is present at $z \sim 2.5$ and then disappears at $z \sim 1.3$ which happens to be the earliest epoch probed by DEEP2 and VVDS.

It has been known for a long time that there are less elliptical galaxies detected at $z = 1$ than at $z = 0$, e.g. Lilly et al.(1995), Cowie et al.(1996), Gabasch et al.(2004), Bell et al.(2004). This implies that a fraction of spiral galaxies have evolved to become elliptical galaxies. Therefore the star formation in these late-type systems must have been quenched to transform them into passive elliptical galaxies. The physical mechanism responsible for this evolution is still a matter of debate but one theory is that AGN feedback is responsible, possibly triggered by galaxy mergers (Hopkins et al.(2007)). The AGN feedback is thought to warm hydrogen gas therefore preventing it from cooling and collapsing to produce new stars. Simulations which include such feedback, e.g. Croton et al.(2006), Bower et al.(2006) provide better agreement with observations such as luminosity or stellar mass functions. Environment may also play an important role through processes such as harassment (Moore et al.(1999)) or ram pressure stripping (Gunn & Gott(1972)) which disrupt gas cooling. Whatever the nature of the mechanism responsible, it primarily acts to turn off star formation in the most massive systems. Therefore the sites of on-going star formation have changed from $z = 1$ to $z = 0$ with star formation found to be taking place in progressively lower mass galaxies, a phenomena referred to as downsizing, e.g. Cowie et al.(1996). This is also found to be reflected in downsizing in halo mass as star formation

moves to lower mass haloes (Heinis et al.(2007)). Couple this with the fact that numerous massive ($M_\star \sim 10^{11}\mathrm{M}_\odot$) galaxies have been detected at $z \sim 2$ (Glazebrook et al.(2004), Cimatti et al.(2004), van Dokkum et al.(2006)), which appears to contradict hierarchical assembly of mass.

Organising these observations (and many more) to form a coherent picture of galaxy formation and evolution and reconciling it with ΛCDM is a challenging task. To fully explain galaxy evolution requires a detailed understanding of physical processes and feedback loops, e.g. star formation, supernovae feedback, galaxy-galaxy interactions, AGN etc. Also understanding more about the typical environments galaxies are found in and how this relates to the underlying dark matter distribution is important. Clustering analyses cannot solve these problems outright but they can provide important insights.

1.3 Analyses of the Clustering of Galaxies

Attempts to quantify the clustering of galaxies are common today with large extra-galactic datasets readily available. Clustering studies are performed for two main purposes. To refine cosmological models by helping to constrain cosmological parameters and to aid our understanding of galaxy environment and evolution. It is the latter of these two which is the focus of this thesis.

1.3.1 Clustering Studies at Low Redshift

The filamentary web-like arrangement of galaxies was first seen in the Lick survey in the 1970's, (Seldner et al.(1977)). The number of galaxies detected was sufficient to make a definite measure of the large scale structure. Davis & Geller(1976) made the first investigation of galaxy clustering as a function of morphologically selected Hubble type using 1942 galaxies from the Uppsala Catalogue (Nilson(1995)). Although their statistics were limited they found the correlation function of elliptical galaxies to have a steeper slope than spiral galaxies. This was an important first step in demonstrating how different types of galaxies cluster differently. Similar studies were subsequently conducted by e.g. Giovanelli et al.(1986), Iovino et al.(1993). In the early 1990's the clustering of infrared selected galaxies was measured by Efstathiou et al.(1990), Georgantopoulos & Shanks(1991), Saunders et al.(1992). Studies using the APM Galaxy Survey provided more clustering results e.g. Maddox et al.(1996), Baugh(1996). Importantly, this enabled

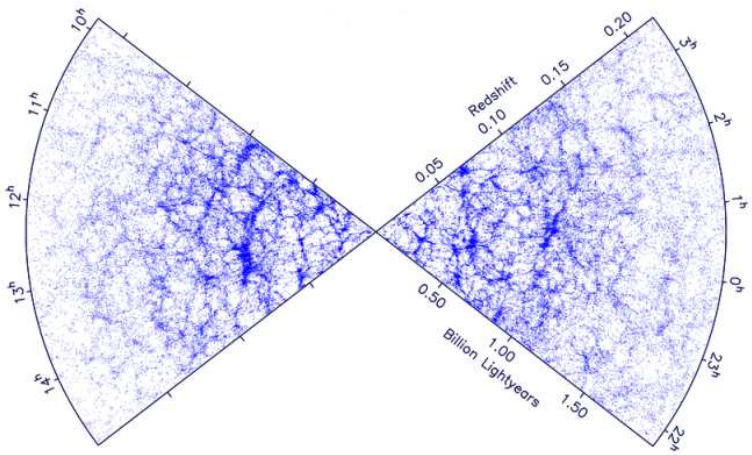


Figure 1.1: The distribution of 63,361 galaxies detected in a 4 degree slice by the 2DFGRS survey. The large scale structure is clearly visible. Plot made by M. Colless (ANU).

clustering as a function of luminosity and morphological type to be investigated in the Stromlo-APM redshift survey (Loveday et al.(1995)) where more luminous galaxies were found to cluster more strongly, a conclusion at odds with some previous studies (see Loveday et al.(1995) and references therein). Loveday et al.(1995) also found early-type galaxies to be much more strongly clustered than late-types. This showed that early-type galaxies reside in denser environments providing a detection of the colour-density relation.

Over the past 10 years dramatic improvements in low redshift clustering analyses were made by a number of surveys. In particular the Sloan Digital Sky Survey (SDSS, York et al.(2000)) and 2 Degree Field Galaxy Redshift Survey (2DFGRS, Colless et al.(2001)) produced a significant number of clustering papers. The extensive galaxy samples detected in these surveys were sufficiently large to be split into subsamples based on galaxy properties and still give statistically significant numbers with which to carry out a clustering analysis.

Clustering as a function of absolute magnitude was studied by Norberg et al.(2001) where they found the clustering amplitude increased with absolute magnitude with a more pronounced increase for $L > L^{\star}$ galaxies. Clustering as a function of colour was investigated by Zehavi et al.(2002) where red galaxies were found to cluster more strongly than blue galaxies and exhibit a steeper correlation function. Similar results were obtained at $z \sim 0.6$ by Phleps et al.(2006). Norberg et al.(2002) looked at the variation of clustering with luminosity and also split their sample into early and late-type galaxies. They found

early-type galaxies to show stronger clustering than late-types in agreement with similar studies e.g. Loveday et al.(1995), Guzzo et al. (1997), Willmer et al.(1998), Firth et al. (2002). They also found the clustering amplitude increased with luminosity at the same rate for both types in agreement with Budavári et al.(2003). Zehavi et al.(2004) also found the clustering strength of blue galaxies to increase with luminosity but the luminosity dependence for red galaxies was less clear. Conversely Shepherd et al. (2001) found the clustering strength of red galaxies to vary with luminosity but blue galaxies to have no dependence on luminosity. This largely depends on the luminosity range probed where a stronger clustering dependence on luminosity is seen for $L > L^*$ galaxies. The clustering of *all* types of galaxies as a function of stellar mass was investigated by Li et al.(2006) where they found more massive galaxies were more strongly clustered than low mass galaxies.

These low redshift studies collectively indicate that different types of galaxies cluster differently. Early-type, redder, older, elliptical galaxies are more strongly clustered than late type, bluer, younger, spiral galaxies with the clustering being largely insensitive to how the two populations are selected. This is not surprising because we know from the colour-density relation that early-type galaxies tend to be found in denser environments whereas late-type galaxies are more frequently found as field galaxies in less dense regions. This shows how clustering studies can provide evidence for the existence of the colour-density relation. For galaxy samples incorporating *all* types of galaxies the clustering strength is seen to increase with luminosity or stellar mass. This is at least partially due to an increase in the ratio of early-to-late-type galaxies, however this is also due to an increase in clustering strength with luminosity, particularly at $L > L^*$ e.g. Norberg et al.(2002), Budavári et al.(2003).

1.3.2 Clustering Studies at High Redshift

Only recently have clustering analyses of galaxy samples selected by spectral type or colour and/or intrinsic galaxy properties become possible at high redshift. Over the past few years there has been a spate of $z \sim 1$ clustering papers from surveys such as VVDS (Meneux et al. 2006, Pollo et al.(2006), Meneux et al.(2008)), DEEP2 (Coil et al.(2006), Coil et al.(2008)), GOODS (Gilli et al.(2007)), MOIRCS (Ichikawa et al.(2007)) and SWIRE (Waddington et al.(2007)) where the large number of galaxies detected allowed statistically significant subsamples to be defined enabling the clustering of galaxies to be analysed as

a function of luminosity, colour, spectral type, stellar mass and/or star formation rate.

Gilli et al.(2007) presented measurements of the clustering strength of $z \sim 0.8$ star forming galaxies with $S_{24\mu\mathrm{m}} > 20\mu$Jy where the clustering strength of their samples were found to increase with star formation rate (SFR). By comparison with the Millennium Simulation it was shown that samples with larger SFR are hosted in progressively more massive haloes. Magliocchetti et al.(2008) looked at the clustering of galaxies detected with $S_{24\mu\mathrm{m}} > 400\mu$Jy in two redshift ranges $0.6 < z < 1.2$ and $z > 1.6$. They found the high z sample to be almost twice as strongly clustered as the low z sample again implying galaxies with higher SFR are found in more massive haloes. These studies provide evidence of downsizing occurring where star formation is predominantly found in lower mass galaxies at low redshift and higher mass galaxies at high redshift.

Pollo et al.(2006) and Coil et al.(2006) measured the clustering as a function of luminosity and found an increase in clustering strength with increasing luminosity. Meneux et al. 2006 investigated the clustering of early- and late-type galaxies finding early-type galaxies to be more clustered than late-type galaxies providing evidence that the colour density relation was in place at $z \sim 1$. Coil et al.(2008) and McCracken et al.(2008) took this a step further and measured the clustering dependence on rest-frame colour and luminosity. Both studies found their red sample to be more strongly clustered than their blue sample. Coil et al.(2008) found the clustering of neither sample to show a variation with luminosity but they did not probe $L > L^*$, where a stronger clustering dependence with luminosity is seen at low redshift. McCracken et al.(2008) found red and blue galaxies to have no dependence on luminosity at $z \sim 0.4$. At $z \sim 0.9$ their red sample showed a steady dependence on luminosity and their blue sample only a weak dependence.

Meneux et al.(2008), Waddington et al.(2007) and Ichikawa et al.(2007) investigated how the clustering of *all* (early and late-type) galaxies detected at $z \sim 1$ depended on the stellar mass content. Meneux et al.(2008) measured the clustering of 3218 galaxies detected by VVDS at $< z >= 0.85$. They split their sample into two mass ranges where they found the clustering strength to almost double from $\mathrm{M}_\star > 10^9\mathrm{M}_\odot$ to $\mathrm{M}_\star > 10^{10.5}\mathrm{M}_\odot$ galaxies. Ichikawa et al.(2007) compared the clustering of low mass galaxies, $10^9 < \mathrm{M}_\star < 10^{10}\mathrm{M}_\odot$, and high mass galaxies, $\mathrm{M}_\star > 10^{10}\mathrm{M}_\odot$, in two redshift ranges $1 < z < 2$ and $2 < z < 4$, detected by MOIRCS in 24.4 arcminutes^2 in the GOODS-N field. They found the more massive galaxies at each redshift are more strongly clustered than the less massive sample. Waddington et al.(2007) measured the clustering of galaxies detected at 3.6μm in the

SWIRE survey. They binned their data into 11 flux limited samples using 3.6μm flux as a tracer of stellar mass where they found the more massive galaxies cluster more strongly.

These high redshift results show early-type or red galaxies were still more clustered than late-type or blue galaxies at $z \sim 1$, indicating the colour-density relation was already in place at these early times. The clustering of early and late-type galaxies was not seen to scale with luminosity, however $L > L^{\star}$ regime was not thoroughly probed. Galaxies selected on star formation rate were found to be more clustered at high redshift than they are at low redshift providing evidence of downsizing. For early and late-type galaxies collectively, more massive galaxies were more clustered but how much of the increase was due to a change in the relative mix of early and late-type galaxies is unclear.

1.3.3 Evolution of Clustering with Redshift

Analyzing the clustering of similar galaxy samples over a range of redshifts enables the evolution of the clustering of galaxies to be quantified and hence helps refine our understanding of galaxy evolution. One of the earliest attempts to analyse the evolution of galaxy clustering was by Warren et al.(1993) who measured the clustering of $z \sim 0.4$ elliptical galaxies and compared it with $z = 0$ results. They found the clustering strength to decrease by about 40% from $z = 0$ to $z = 0.4$. Other early clustering work out to $z \sim 1$ by Le Fevre et al. (1996), Carlberg et al. (1997) and Hogg et al. (2000) also found the clustering strength to decrease with increasing redshift.

Over the past few years data sets spanning large redshift ranges have become more common. This has lead to a number of studies which looked at the evolution of galaxy clustering strengths. Waddington et al.(2007) found the clustering strength to be approximately constant at $z < 0.5$ and to decrease over $0.5 < z < 1$. However they used 11 samples with increasing 3.6μm flux limits. Hence the average mass of each sample decreased with increasing redshift. They were also mixing early and late-type galaxies. Therefore it was difficult to determine how much of the change in clustering strength was a real physical trend and how much was due to a change in the mean mass of the samples and/or the mix of early and late-type galaxies. de la Torre et al.(2007) also investigated the clustering of SWIRE galaxies selected on 3.6μm flux over $0.2 < z < 2.0$. They used the photometric redshifts of Ilbert et al.(2006) to more accurately constrain the redshift distribution. They found the clustering strength to be constant with redshift. However, they used the same flux cut for all redshift bins and hence selected progressively more massive

galaxies with increasing redshift, as well as mixing early and late-type galaxies. Le Fèvre et al.(2005) found the clustering strength to increase slightly with redshift however they were selecting progressively more luminous galaxies with increasing redshift and mixing types. By comparison with the clustering of similar luminosity galaxies at $z = 0$ detected by 2DFGRS they concluded the clustering strength decreased by a factor of ~ 2.5 over $0 < z < 2.0$. Meneux et al.(2008) compared their correlation functions with those of Li et al.(2006) at $z \sim 0.1$ and found no evolution in the correlation function for their high mass sample but their low mass sample evolved implying clustering evolution has been strongest from $z \sim 1$ to the present day for low mass galaxies.

The vast majority of clustering analyses presented in the literature are not suitable for investigating the evolution of clustering. Studies which have selected galaxies based on apparent magnitude (or flux) make it difficult to investigate the evolution of clustering because galaxies selected at low and high redshifts have different luminosites or stellar masses, which we have seen cluster differently. Also attempting to quantify the evolution of clustering by comparisons across surveys is dangerous due to different selection effects inherent to each survey. Studies which have not separated elliptical/early-type/red galaxies from spiral/late-type/blue galaxies cannot be relied upon because it is impossible to separate the clustering signal from each type of galaxy and we do not know how the relative mix of each type varies with redshift.

To reliably analyse variations in the clustering strength with redshift requires galaxies of a specific type to be selected based on intrinsic properties. One study which has performed such an analysis is McCracken et al.(2008). They selected galaxies based on absolute B-band magnitude and rest-frame colour over $0.2 < z < 1.0$. They found the clustering strength of blue galaxies to be invariant with redshift (and luminosity). The clustering of their most luminous red galaxies did not change with redshift but the less luminous red galaxies showed a decrease with redshift.

Clustering analyses where early and late-type galaxies are selected based on intrinsic properties over a large redshift range are few and far between. This is unfortunate as such studies can provide important insights into galaxy evolution. However it is not surprising considering the technical challenges involved in obtaining redshifts for large samples of galaxies out to high redshift. Photometric redshifts provide a way to obtain such a dataset. To date a clustering analysis of early and late-type galaxies selected on stellar mass content or star formation rate across a large redshift range has not been

performed. This thesis performs such an analysis using Rowan-Robinson et al.(2008)'s photometric redshift catalogue for the Spitzer Wide area InfraRed Extragalactic (SWIRE) survey, where multiple fields and large volumes are probed to high redshift.

1.4 Summary of the Thesis

The majority of this thesis is dedicated to measuring the clustering of early and late-type galaxies detected as part of the SWIRE survey as a function of stellar mass content and star formation rate over $0 < z < 1.5$. An analysis of the clustering of dark matter haloes and synthetic galaxies in the Millennium Simulation is also presented. Finally, a search for population III stars in the Spitzer dark field is conducted.

The SWIRE photometric redshift catalogue of Rowan-Robinson et al.(2008) is utilized to measure the angular correlation function of galaxies as a function of spectral type, stellar mass content and star formation rate over $0 < z < 1.5$. The large size of the catalogue enables the variation of clustering with these properties to be thoroughly investigated. Hence this catalogue provides an excellent opportunity for clustering measurements to be used to constrain theories of galaxy evolution. This data set also provides us with the opportunity to investigate the clustering of early and late-type galaxies to high stellar masses $\log(M_\star/M_\odot) \geq 11.2$ which has never before been probed at $z > 0.4$. It also enables the evolution of the clustering of galaxies to be measured out to $z \sim 1.5$ extending beyond $z \sim 1$ previously attained by McCracken et al.(2008).

Chapter 2 provides definitions of the correlation function and outlines it's application to quantifying the clustering of galaxies. Other methods for measuring clustering are discussed. Chapter 3 gives a detailed description of the SWIRE photometric redshift catalogue and a thorough investigation of it's complicated selection effects. Criteria are determined to constrain the selection effects so the catalogues can reliably be used to carry out a clustering analysis. In Chapter 4 the Millennium Simulation is introduced and the clustering of simulated dark matter haloes is measured over a range of redshifts. Also the clustering of synthetic galaxies, from a semi-analytic model, selected on stellar mass is measured over $0 < z < 1.5$. Chapter 5 outlines how the clustering of galaxies in the photometric redshift catalogue was measured. The catalogues were binned by redshift and stellar mass producing stellar mass-redshift *cells* denoted $(M_\star, z)$ and separately by star formation rate and redshift producing (SFR, z) cells. The angular correlation function of

the galaxies in each cell was measured. The results for the clustering analyses are presented aswell as comparisons to other surveys. The clustering of the SWIRE galaxies is compared to that of the Millennium Simulation dark matter haloes to ascertain typical parent halo masses. Chapter 6 outlines a search for population III stars using the Spitzer dark field. The dark field contains repeat infrared imaging at 2-3 week intervals of unprecedented depth making it ideal for a search for population III stars. In Chapter 7 conclusions from the clustering analyses and population III star search are summarized and a short discussion on how this work can be taken forward is given.

Chapter 2

Measuring Large Scale Structure

2.1 Introduction

In this chapter the correlation function is introduced as a statistical method used to measure the clustering of galaxies. The evolution of the correlation function with redshift is parameterized. A simple estimator for calculating the correlation function is derived. Other methods for doing the pair counting are discussed including counts-in-cells and kd-tree methods.

2.2 The Correlation Function

The large scale structure of the Universe is a remnant of the density variations of the dark matter density field in the early Universe. As dark matter cannot be seen or directly detected the large scale structure of the Universe can be mapped out by using galaxies as tracers of the peaks in the density field.

By considering an angular distribution of galaxies on the sky with mean density η, the probability of finding a galaxy located in a solid angle $d\Omega$ is (Peebles 1984),

$$dP = \eta \, d\Omega \tag{2.1}$$

The mean number of objects in a finite area Ω is

$$\langle N \rangle = \eta\Omega \tag{2.2}$$

The distribution of galaxies is often quantified by measuring their *two point auto-correlation function*. For an astronomical survey traditionally the *angular correlation function*, $\omega(\theta)$,

is calculated as this does not require knowledge of the redshifts of the galaxies. $\omega(\theta)$ is defined as the excess probablility of finding one galaxy in solid angle $d\Omega_1$ and another galaxy in $d\Omega_2$ separated on the sky by an angle θ

$$dP(\theta) = N_\Omega^2[1 + \omega(\theta)]d\Omega_1 d\Omega_2 \tag{2.3}$$

where N_Ω is the mean number density of sources (per steradian) in the survey (e.g. Phillipps et al.(1978)). The correlation function also manifests itself in the *spatial* version. This is particularly useful for measuring the clustering of simulated data or astronomical data where accurate depth information is known. The two-point spatial correlation function $\xi(r)$ at redshift z, gives the excess probablility of finding a neighbour at a distance r from a given source. The probability of finding one source in proper volume elements dV_1 and a second source in dV_2 is

$$dP(r, z) = n^2(z)[1 + \xi(r, z)]dV_1 dV_2 \tag{2.4}$$

Phillipps et al.(1978), where $n(z)$ is the mean number density of sources per unit of proper volume. Both forms are often modelled as a power law with the spatial version taking on the form

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \tag{2.5}$$

where r_0 is the spatial clustering length (which shall be referred to frequently throughout this thesis as the *clustering strength*) and γ parameterizes the scale dependence. r_0 is not a physical lengthscale in the spacial distribution of galaxies but is the length at which $\xi(r)$ is unity (i.e. the chance of finding a galaxy at a distance r_0 from another galaxy is twice the Poissonian chance). The angular version is parameterized as

$$\omega(\theta) = A\theta^{1-\gamma} \tag{2.6}$$

where A is the amplitude of the correlation function. $\omega(\theta)$ is a function of $|\theta|$ only. If $\omega(\theta)$ is zero, the distribution of galaxies is a Poisson distribution and is considered to be unclustered. Whereas $\omega(\theta) > 0$ indicates clustering and $\omega(\theta) < 0$ indicates anti-clustering, i.e galaxies tend to avoid each other.

2.2.1 Evolution of the Correlation Function

The evolution of the correlation function, in comoving coordinates, may be written as

$$\xi(r, z) = \left(\frac{r}{r_0}\right)^{-\gamma} (1 + z)^{\gamma-(3+\epsilon)} \tag{2.7}$$

Overzier et al.(2003) where r_0 is the comoving correlation length measured at $z = 0$ and ϵ parameterizes the evolution with redshift. The comoving correlation length at redshift z is given by

$$r_0(z) = r_0(0)\ (1+z)^{1-(3+\epsilon)/\gamma} \tag{2.8}$$

where $r_0(z)$ is the correlation length at the redshift in question and $r_0(0)$ is the correlation length at $z = 0$.

2.2.1.1 Interpretation of ϵ

In Equation 2.8 the factor of $(1 + z)^{-3}$ allows for a change in the mean density due to expansion. ϵ determines how the correlation function evolves over time (Phillipps et al.(1978), Overzier et al.(2003)). Three qualitative scenarios are commonly considered. $\epsilon = 0$ corresponds to the *stable* clustering model where the clustering strength increases as we approach $z = 0$ as clusters remain fixed in proper coordinates and the background mass distribution dilutes. The comoving clustering model is where $\epsilon = \gamma - 3$ and haloes expand with the Universe, therefore their clustering remains fixed over time unless they are still growing. We can see this is so by substituting $\epsilon = \gamma - 3$ into Equation 2.7 where the exponent of $(1+z)$ reduces to zero. This is the value of ϵ used in Limber's equation in Chapter 5 to determine the comoving correlation length, r_0. The third model sets $\epsilon = \gamma - 1$ where clustering grows over time in accordance with linear perturbation theory, (Carlberg et al.(2000)).

2.3 Correlation Function Estimators

Correlation function errors are known to be the sum of two physically separate contributions, namely Poisson shot noise (due to the sampling of the underlying continuous density field with a finite number of galaxies) and sample variance (due to the fact that only a finite spatial volume is probed), Swanson et al.(2008). On large scales sample variance dominates due to fewer pairs of galaxies at these separations. Large and/or multiple fields can be used to help overcome these problems.

There are a number of estimators to choose from when measuring the correlation function. A quick derivation of the simplest estimator (Equation 2.12), adapted from Sicotte & Peebles(1995) is given below. It follows from Equation 2.3, by considering a catalogue of n_d data sources covering a solid angle Ω, that the mean number of sources at a distance

$\theta \pm \Delta\theta$ from a source chosen at random is

$$(n_d - 1)(1 + \omega(\theta))\frac{\langle \delta\Omega \rangle}{\Omega} \tag{2.9}$$

where $\langle \delta\Omega \rangle$ is the mean solid angle about a randomly chosen source. The total number of data-data pairs of sources with separations in the interval $\theta \pm \Delta\theta$ is then given by

$$DD(\theta) = \frac{n_d}{2}(n_d - 1)(1 + \omega(\theta))\frac{\langle \delta\Omega \rangle}{\Omega} \tag{2.10}$$

$DD(\theta)$ can be measured directly from the catalogue, and when combined with an estimate of $\langle \delta\Omega \rangle / \Omega$, can give an estimate of $\omega(\theta)$. For surveys with complicated geometries, it is more practical to calculate $\langle \delta\Omega \rangle / \Omega$ using a catalogue of randomly distributed points covering the same area as the survey. For a random catalogue of n_r random sources the number of pairs of random sources with separations between $\theta \pm \Delta\theta$ is given by

$$RR(\theta) = \frac{n_r}{2}(n_r - 1)\frac{\langle \delta\Omega_r \rangle}{\Omega} \tag{2.11}$$

This quantity can be measured and a value for $\langle \delta\Omega_r \rangle / \Omega$ can be obtained which is approximately equal to $\langle \delta\Omega \rangle / \Omega$. A simple estimator of $\omega(\theta)$ is therefore given by

$$\omega(\theta) = \frac{DD(\theta)}{RR(\theta)} - 1. \tag{2.12}$$

In practice $n_r \gg n_d$ so that the estimate of $\omega(\theta)$ is not limited by statistical errors. Throughout this thesis typically $n_r = 500,000$ was used when measuring the clustering of a catalogue containing $n_d \sim 1000$ data sources (see Chapter 5). An improvement on the estimate of $\langle \delta\Omega \rangle / \Omega$ can be obtained by counting the number of data-random pairs, DR(θ). Using DR instead of RR enables the mean solid angle around data sources $[\langle \delta\Omega \rangle / \Omega]$ to be measured as opposed to the mean solid angle around random sources $[\langle \delta\Omega_r \rangle / \Omega\ ]$. The measurements of RR and DR correct for the *edge effects*, i.e. they estimate the mean solid angle about points in the area of the survey for a given θ.

The most commonly used estimator, which is adopted throughout this thesis, is the Landy & Szalay(1993) estimator

$$\omega(\theta) = \frac{DD - 2DR + RR}{RR} \tag{2.13}$$

whose variance is effectively poisson, see Landy & Szalay(1993) for a full description. DD, DR and RR are all a function of θ and normalised by the respective total number of pairs over all scales

$$DD = \frac{dd}{\frac{1}{2}n_d(n_d - 1)} \tag{2.14}$$

$$DR = \frac{dr}{n_d\ n_r} \tag{2.15}$$

$$RR = \frac{rr}{\frac{1}{2}n_r(n_r - 1)} \tag{2.16}$$

where n_d is the number of data galaxies and n_r is the number of random galaxies.

2.4 The Integral Constraint

When measuring the clustering of a sample of galaxies we are restricted to a bound area of sky and hence lack an estimate of the average global density for the type of galaxies in question. This leads to the measured $\omega(\theta)$ being different from the true $\omega_t(\theta)$ because we assume the density of galaxies in our field to be representative of the global value. This can be corrected for in measurements of the angular correlation function by introducing a constant commonly known as the integral constraint. This can be calculated analytically,

$$I_c = \frac{1}{\Omega^2}\int\int \omega(\theta) d\Omega_1 d\Omega_2 \tag{2.17}$$

where Ω is the area of the field. However there are two ways to calculate the integral constraint numerically, both using the unnormalised random counts. The first method is to calculate

$$\delta = \frac{\Sigma_i N_{rr}(\theta_i)}{\Sigma_i N_{rr}(\theta_i)(1 + \omega(\theta_i))} \tag{2.18}$$

and divide $\omega(\theta)$ by δ. This is an iterative process but the *true* value of $\omega(\theta)$ converges after 3 or 4 iterations. The other method, following Roche et al. 1999, is to model the correlation function as

$$\omega_t(\theta) = A(\theta^{1-\gamma} - C). \tag{2.19}$$

where AC is the integral constraint. C can be calculated numerically using

$$C = \frac{\sum_j N_{rr}(\theta_j)\theta_j^{1-\gamma}}{\sum_j N_{rr}(\theta_j)} \tag{2.20}$$

Throughout this thesis we adopt the second method.

2.5 Limber's Equation

The angular correlation function is a measure of the projection of the galaxies on the sky, however galaxies which appear close together may actually be separated by a large distance along the radial direction. $\omega(\theta)$ is the convolution of the redshift distribution and

the spatial correlation function. The spatial correlation length, r_0, can thus be calculated from the amplitude of the angular correlation function, A, using Equation 2.21, commonly known as Limber's equation, Peebles 1953,

$$A = \frac{r_0(0)^\gamma \ f}{c} \frac{\int_{-\infty}^{\infty} H(z) \ (1+z)^{-(2+\epsilon)} \ D_A^{1-\gamma} \ (dN/dz)^2 \ dz}{\left[\int_0^{\infty} (dN/dz) \ dz\right]^2} \tag{2.21}$$

where D_A is the angular diameter distance and dN/dz is the redshift distribution which implicitly includes the selection function, $\phi(z)$,

$$\frac{dN}{dz} = n(z) \ \phi(z) \ c \ H^{-1} D_A^2 \ (1+z)^{-1} \tag{2.22}$$

$H(z)$ is the Hubble parameter given by,

$$H(z) = H_0 \sqrt{\Omega_M (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda} \tag{2.23}$$

and f is a constant given by,

$$f = \frac{\sqrt{\pi} \ \Gamma([\gamma - 1]/2)}{\Gamma(\gamma/2)} \tag{2.24}$$

where Γ is the standard gamma function.

2.6 Other Methods for Calculating the Correlation Function

The correlation function is the most commonly used method to quantify the clustering of galaxies, usually done through pair counting. The simplest way to calculate the separation of n galaxies requires $n(n-1)$ computations. When using large data sets this can become a time consuming process, especially if a large random catalogue is used, and even if a powerful computer is utilized. There are other methods, apart from pair counting, used to calculate the correlation function which do have their advantages in certain situations. A common method to reduce the number of calculations required and hence speed up the calculation is to use a counts-in-cells approach.

2.6.1 Counts-in-Cells

A counts-in-cells method works by gridding up the spatial planes of the data into an array of bins. The pair counting is then performed on the bins instead of the data sources thereby greatly reducing the number of computations required. This method effectively smoothes the data on scales corresponding to the bin size.

The following equations outline how to calculate $\xi(r)$ using a counts-in-cells method for synthetic data. Consider a simulation cube with periodic boundary conditions containing N_d data sources each with x,y and z cartesian coordinates. This is applicable for simulations such as the Millenium Simulation (see Chapter 4). The simulation cube is split into n_c smaller sub-cubes and the number of sources in each sub-cube, n, is counted. The fluctuation of sources in the i^{th} sub-cube, δ_i, and j^{th} sub-cube δ_j are given by

$$\delta_i = \frac{n_i - \bar{n}}{\bar{n}} \tag{2.25}$$

$$\delta_j = \frac{n_j - \bar{n}}{\bar{n}} \tag{2.26}$$

where $\bar{n}$ is the mean number of galaxies in each sub-cube and n_i and n_j are the number of sources in i^{th} and j^{th} sub-cubes, respectively. For every pair of sub-cubes separated by a distance r, the correlation function can be calculated using,

$$\xi(r) = \langle \delta_i \delta_j \rangle \tag{2.27}$$

where the average is over all sub-cubes. Substituting Equations 2.25 and 2.26 into Equation 2.27 gives,

$$\xi(r) = \left\langle \frac{(n_i - \bar{n})(n_j - \bar{n})}{\bar{n}^2} \right\rangle \tag{2.28}$$

We developed an IDL code to calculate the spatial correlation function using Equation 2.28. The code was designed to calculate $\xi(r)$ one scale r at a time. δ was calculated for every cell producing a three dimensional array where each element corresponded to δ for a specific sub-cube, lets call this array δ_1. δ_1 was then translated (using the IDL SHIFT function) by a distance equal to r, creating array δ_2. Note the data used had periodic boundary conditions. The product of these two arrays was then calculated and stored. δ_1 was then translated again by r along a different direction producing δ_3 of which the product with δ_1 was taken. If we consider every sub-cube to be a distance r from n_a other sub-cubes then in total δ_1 was translated n_a times. For all $i-j$ pairs of sub-cubes separated by a distance r the correlation function could be calculated using,

$$\xi(r) = \frac{1}{n_a n_c} \sum_{j}^{n_a} \sum_{i,jpair}^{n_c} \delta_i \delta_j \tag{2.29}$$

This could be repeated for all desired scales. We calculated the Poisson error on the correlation function by initially considering the error on the counts in each sub-cube,

$$\sigma_n = \sqrt{n} \tag{2.30}$$

Hence the error on $\bar{n}$ is given by,

$$\sigma_{\bar{n}} = \sqrt{\frac{\bar{n}}{n_c}} \tag{2.31}$$

which gives an error on δ_i given by,

$$\sigma_{\delta_i} = \sqrt{\frac{n_i\bar{n} + (n_i^2/n_c)}{\bar{n}^3}} \tag{2.32}$$

Considering Equation 2.27, this leads to an error on the product of $\delta_i\delta_j$ given by,

$$\sigma_{\delta_i\delta_j} = \delta_i\delta_j\sqrt{\left(\frac{\sigma_{\delta_i}}{\delta_i}\right)^2 + \left(\frac{\sigma_{\delta_j}}{\delta_j}\right)^2} \tag{2.33}$$

and hence an error on the correlation function given by

$$\sigma_{\xi(r)} = \frac{1}{n_a n_c}\sqrt{\sum_{j}^{n_{an}}\sum_{i,jpair}^{n_c}\sigma^2_{\delta_i\delta_j}} \tag{2.34}$$

for all $i-j$ sub-cubes separated by a distance r.

The code was tested using a synthetic data set. A catalogue of 100,000 sources was generated with a Possion distribution in the x, y, z planes. The sources covered the following ranges; $0 < x < 65$, $0 < y < 65$, $0 < z < 65$. Figure 2.1 shows the spatial correlation function for this data set measured using the counts-in-cells method where the sub-cube size was set at 2. $\xi(r)$ is consistent with zero at all scales as expected. Figure 2.2 shows an example of a correlation function measured using the counts-in-cells method on a dark matter halo catalogue from the Millennium Simulation.

Although counts-in-cells methods are fast, the error on $\xi(r)$ tends to be larger and scales smaller than the sub-cube size cannot be probed. A counts-in-cells method is useful when an approximate answer is required quickly.

2.6.2 Kd-Tree Methods

A fast method for performing the pair counting uses a k-dimensional tree (kd-tree). A kd-tree is a space-partitioning data structure used to organize points in a k-dimensional space, Gao et al.(2008). Put more simply a kd-tree groups data points in close proximity to each other into bounding boxes which are organized hierarcially. Whenever the tree is queried requesting a list of all points in a neighbourhood, the query can be executed quickly without needing to visit every single data point.

Constructing a tree usually begins at the *root* level (i.e. all of the data) and branches out from there. Consider the angular distribution of a group of galaxies. The idea is to

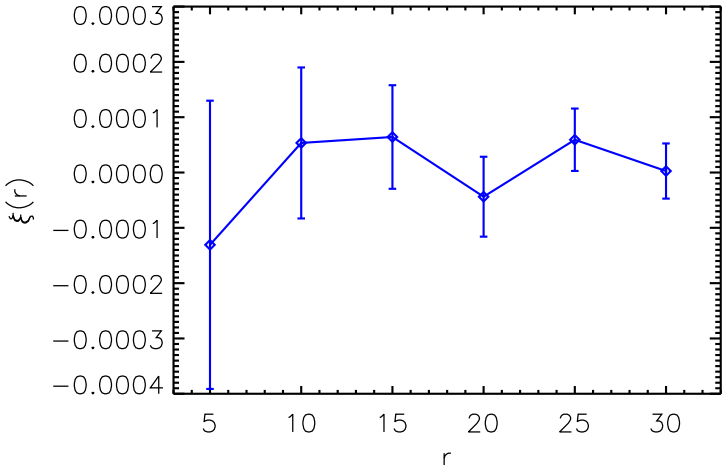


Figure 2.1: The spatial correlation function measured using the counts-in-cells technique using a Poisson distribution of 100,000 sources. The size of each sub-cube was 2 and the full spatial range was 65. The errors are Poisson errors.

recursively partition the angular distribution of galaxies by separating along the widest dimension. Each time the data is split at the centre of the widest dimension creating two *child* nodes which divide their parent's data between them. This process is repeated until the widest dimension of a bounding box is less than some predetermined threshold. This is then known as a *leaf* node.

Once the tree has been made, queries regarding the angular distribution of galaxies can be executed quickly without having to measure the separation of every pair. The correlation function is treated as a *range finding* problem. Each galaxy is taken in turn to be a *query point* and then a range search of the kd-tree is performed to find all other points within a distance r of the query point. The search is a depth-first traversal of the kd-tree, where the minimum distance, d_{min}, between the query point and the bounding box surrounding the current node is measured. If $d_{min} > r$ there is no need to visit the node's children and computation is saved. Also the kd-tree often stores extra information known as *cached sufficient statistics* (Moore & Lee (1998)). For the purpose of the correlation function, for each node the number of galaxies within the bounding box can be counted and stored. This is particularly useful when $d_{min} < r$ because then all galaxies within the bounding box are within range of the query point and the separation of each galaxy with the query point does not have to be evaluated.

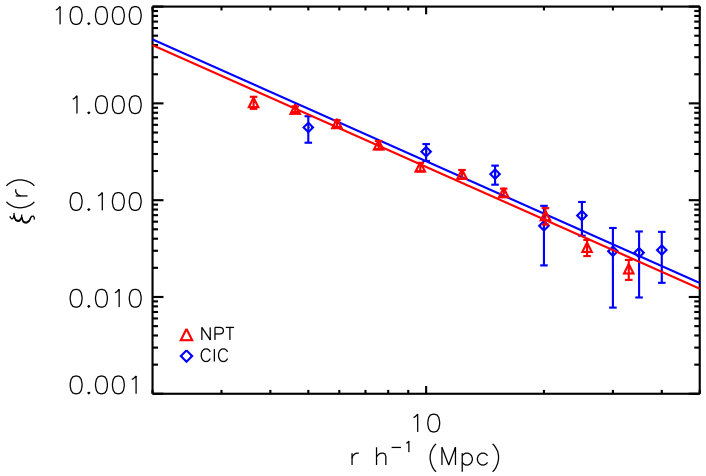


Figure 2.2: The spatial correlation function measured using the counts-in-cells (CIC, blue diamonds) and NPT (red triangles) methods. Agreement between the two methods is good.

A pair-counting code utilizing kd-tree methods, the NPT code, was provided by The Auton Lab at Carnegie Mellon University's School of Computer Science (http://www.autonlab.org) and we thank Andrew Moore et al. for making this software public. Due to the accuracy and speed of the code it's use is adopted throughout this thesis. See Gray et al.(2001) for a full description of the code.

Figure 2.2 shows the spatial correlation function measured for a test data set using the counts-in-cells and NPT code. The data used was a halo catalogue from the Milennium Simulation at $z = 0.1$ containing haloes of mass $\log(M_{halo}/\mathrm{M}_{\odot}) = 12$. The errors for the counts-in-cells and NPT methods are Poisson errors and jackknife errors, respectively. The agreement between the two methods is good with the counts-in-cells method showing larger errors.

2.7 Summary

The angular correlation function was introduced which was parameterized as a power law (Equation 2.6). This is the statistical method used to calculate the clustering of galaxies in Chapter 5. An overview of the NPT pair counting code was given which uses kd-trees to improve the time taken to do the pair counting. Limber's equation was also defined which is used to calculate the spatial clustering strength, r_0, from the amplitude of the angular

correlation function in Chapter 5.

Chapter 3

The SWIRE Photometric Redshift Catalogue and Selection Effects

3.1 Introduction

In this chapter a short introduction to infrared astronomy is given along with a description of the SWIRE survey. The SWIRE photometric redshift catalogue (Rowan-Robinson et al.(2008)) is introduced which is integral to the research in this thesis. The complicated selection effects inherent to the catalogue, arising from the combination of infrared and optical data, are thoroughly investigated. This involves ascertaining the completeness of the catalogue using optical and infrared completeness curves. Also the quality of the photometric redshifts are investigated. Criteria are defined to provide a data set suitable to use for a clustering analysis.

3.2 Infrared Astronomy

Prior to 1930 it was commonly accepted that the interstellar medium was a perfect vacuum. In 1930 Robert Trumpler measured the distance to about 100 star clusters using two independent methods (Trumpler(1930)). In one method he used the spectral type of a star in a cluster to estimate the absolute magnitude. Then by measuring the apparent magnitude he could estimate the distance to the cluster. The second method involved estimating the distance to the clusters by measuring their angular diameter distance. He found the two methods did not agree and that the affect was more pronounced the further

away the cluster was. He concluded that space is filled with dust causing the distant clusters to appear less bright than their nearby counterparts.

We now know that post-main sequence stars that have entered the giant phase of their evolution are an important source of dust grains in galaxies (e.g. Lugaro et al.(2003), Draine(2003)). Large grains start with silicate particles forming in the atmospheres of cool stars, and carbon grains in the atmospheres of cool carbon stars. Physical processes such as stellar winds and supernovae transport these grains into the interstellar medium which can become an important source of material for subsequent star formation. Dust grains absorb optical and ultraviolet light from surrounding stars causing the dust to heat up and re-emit the energy at mid-infrared to sub-mm wavelengths.

The infrared part of the electromagnetic spectrum extends from $1-1000\mu$m. Different parts of the infrared spectrum provide us with different information about galaxies. At near-infrared wavelengths ($\sim 0.8-8.0\mu$m) dust is almost transparent. We see light primarily from cooler red giant stars and low mass red dwarfs, which comprise most of the stellar mass content of a galaxy. Therefore near-infrared wavelengths are good for estimating the total stellar mass of a galaxy. For example K-band (2.2μ m) surveys allow galaxies to be selected in this way up to $z \sim 1.5$. Optical and UV light from stars is absorbed by dust, which is warmed to $140-740K$ and re-emitted in the mid-infrared regime ($8.0-20\mu$m). At far-infrared wavelengths ($20-300\mu$m) emissions from cold dust ($< 140K$) are detected which are the sites of star formation. Therefore this regime is a good tracer of ongoing star-formation activity in a galaxy.

To provide a physical underpinning to infrared observations, models of amorphous silicate grains and carbonaceous grains have been developed which can successfully reproduce the observed amounts of extinction due to dust given an appropriate distribution in the size of the grains (e.g. Weingartner & Draine(2001), Li & Draine(2001)). Strong interstellar emission features observed at 3.3μm, 6.2μm, 7.7μm, 8.6μm, 11.3μm and 12.7μm were originally referred to as Unidentified InfraRed (UIR) bands. Dust models first developed by Leger & Puget(1984) and Allamandola et al.(1985) and subsequently improved upon by Desert et al.(1990) and Li & Draine(2001) identified these bands to originate from vibrational modes of Polycyclic Aromatic Hydrocarbons (PAHs). The stretching of the $C-H$ bond produces the 3.3μm feature, $C-H$ bending modes produce the in-plane 8.6μm and out-of-plane 11.3μm features, and $C-C$ stretching and bending modes produce the emission features at 6.2μm, 7.7μm, 8.6μm, 11.3μm and 12.7μm.

3.3 Observing in the Infrared

3.3.1 IRAS

The InfraRed Astronomical Satellite (IRAS) (Neugebauer et al.(1984)) was launched on the 25th January 1983. It conducted an all-sky survey at 12μm, 25μm, 60μm and 100μm during it's 10 month mission. Some of it's most important achievements included the discovery of galaxies with large IR luminosities, namely luminous infrared galaxies (LIRGs) with $L_{FIR}/\mathrm{L}_{\odot} > 10^{11}$, ultra-luminous infrared galaxies (ULIRGs) $L_{FIR}/\mathrm{L}_{\odot} > 10^{12}$ and hyper-luminous infrared galaxies (HLIRGs) $L_{FIR}/\mathrm{L}_{\odot} > 10^{13}$. IRAS also detected $\sim 75,000$ starburst galaxies indicating they are not as rare as previously thought.

3.3.2 ISO

The Infrared Space Observatory (ISO) (Kessler et al.(1992)) was launched on the 17th November 1995. ISO imaged the sky at 6.7μm, 12μm, 15μm, 60μm, 90μm, 135μm, 175μm and 180μm, thereby probing to longer and shorter wavelengths than IRAS with a sensitivity 1000 times better. ISO undertook numerous extra-galactic surveys, see Table 1 in Oliver et al.(2000b) for an overview. One of these surveys was the European Large Area ISO Survey (ELAIS, Oliver et al.(2000a)) which was comprised of four main fields N1, N2, N3 in the northern hemisphere and S1 in the southern hemisphere. N1, N2 and S1 would later be re-observed as part of the SWIRE survey, see Section 3.4.

ISO showed through the decrease in the number counts of LIRGs and ULIRGs from $z \sim 1$ to $z = 0$ that such galaxies were more common at high redshift, see Elbaz(2005) and references therein. Through investigations of the central regions of 15 ULIRGS Genzel et al.(1998) began to distinguish between starbursts and active galactic nuclei (AGN) as the source of luminous IR emission in such galaxies.

3.3.3 The Spitzer Space Telescope

The launch of the Spitzer Space Telescope (Werner et al.(2004)) on the 25th August 2003 marked the beginning of a new era for observations of the Universe at infrared (IR) wavelengths. It was launched into an Earth-trailing orbit, where it loses ground at a rate of 0.1 AU per year, keeping the infrared glare from the Earth sufficiently small in it's field of view. Spitzer is equipped with three instruments; the InfraRed Array Camera (IRAC, Fazio et al.(2004), see Section 3.3.3.1), the Multiband Imaging Photometer for Spitzer

(MIPS, Rieke & MIPS Team(2004), see Section 3.3.3.2) and the InfraRed Spectrograph (IRS, Houck et al.(2004)). Spitzer was initially intended to be operational for two and half years, with the possibility of extending to five. In May 2009 it entered the warm phase of it's mission as it's cryogenic tanks became empty. It will continue observing with the IRAC camera for one more year. The limiting factor on it's life being funding and not technical or mechanical failure.

3.3.3.1 IRAC

IRAC is an imaging instrument onboard Spitzer designed to observe simultaneously in four bands centred on 3.6μm, 4.5μm, 5.8μm and 8.0μm with each having a field-of-view of $5.12' \times 5.12'$. The detector array for each channel is comprised of 256×256 pixels. The wavelengths probe the near-to-mid infrared.

3.3.3.2 MIPS

MIPS has three detectors centred on 24μm, 70μm and 160μm designed to operate simultaneously. The detector array for the 24μm channel is 128×128 pixels, for 70μm 32×32 pixels and at 160μm is 2×20 pixels. These wavelengths span the mid-to-far infrared. Approximately 40% of the emission from starburst galaxies is in the range $8 - 40\mu$m. The 24μm channel is sensitive to cold dust emission and hence is well placed for estimating the amount of on-going star formation in a galaxy.

3.4 The SWIRE Survey

The Spitzer Wide-area InfraRed Extragalactic Survey (SWIRE, Lonsdale et al.(2003)) is the largest of the legacy surveys on the Spitzer Space Telescope. SWIRE is a wide-area imaging survey designed to trace the evolution of dusty, star-forming galaxies, evolved stellar populations, and active galactic nuclei (AGN) as a function of environment, from redshifts $z \sim 3$ to the current epoch. The SWIRE survey began making observations in 2003 and was completed in 2006. It covered a total of 49 square degrees in six different fields; ELAIS-N1, ELAIS-N2, the Lockman hole, ELAIS-S1, XMM and Chandra Deep Field South (CDFS). The six fields were chosen to be in regions of the sky that contained the most ancillary data from other wavebands and to be at high galactic latitudes where contamination from cirrus is negligible, see Figure 3.1 and Table 3.1. Each field was

Table 3.1: Information about the six SWIRE fields. The columns are as follows; field name, the location of the centre of the field and the area of the field.

Field	Field Centre J2000	Total Area deg^2
ELAIS-N1	$16^{\rm h}\,11^{\rm m}\,00^{\rm s} + 55^\circ\,00'\,00''$	9.0
ELAIS-N2	$16^{\rm h}\,36^{\rm m}\,48^{\rm s} + 41^\circ\,01'\,45''$	4.2
ELAIS-S1	$00^{\rm h}\,38^{\rm m}\,30^{\rm s} - 44^\circ\,00'\,00''$	6.9
Lockman	$10^{\rm h}\,45^{\rm m}\,00^{\rm s} + 58^\circ\,00'\,00''$	11.0
CDFS	$03^{\rm h}\,32^{\rm m}\,00^{\rm s} - 28^\circ\,16'\,00''$	7.8
XMM	$02^{\rm h}\,21^{\rm m}\,20^{\rm s} - 04^\circ\,30'\,00''$	9.1

imaged using the four IRAC bands centred on 3.6μm, 4.5μm, 5.8μm and 8.0μm and the three MIPS bands at 24μm, 70μm and 160μm. The area and depth of SWIRE combine to produce a survey of comoving volume 0.2 h^{-3}Gpc^3 over $0 \leq z \leq 2$, and spatial scales of $\sim 100\ h^{-1}$Mpc at $z \sim 1$. For the clustering analysis contained in this thesis we work exclusively in ELAIS-N1 and ELAIS-N2.

3.4.1 SWIRE Coverage Maps

For each of the SWIRE fields a coverage map was produced which recorded the number of independent images contributing to each pixel, usually at least 4. These coverage maps also took into account the complex dithering pattern produced by the scan strategy of the satellite and any missing data due to cosmic ray rejection. The coverage map is therefore proportional to the integration time. Pixels with a mean coverage less than 2.95 were excluded from the coverage map (this allowed for up to one of the images in the full coverage areas to be flagged and rejected due to a cosmic ray). It is assumed the noise in the image, σ, is inversely proportional to the square root of the integration time t

$$\sigma \propto \frac{1}{\sqrt{t}} \tag{3.1}$$

and so varies with coverage κ, also as

$$\sigma \propto \frac{1}{\sqrt{\kappa}} \tag{3.2}$$

Figures 3.2(a) to 3.2(e) show the IRAC 3.6μm coverage maps (the 4.5μm, 5.8μm and 8μm look very similar) and Figures 3.3, 3.4, 3.5 a) to e) show the MIPS 24μm, 70μm and

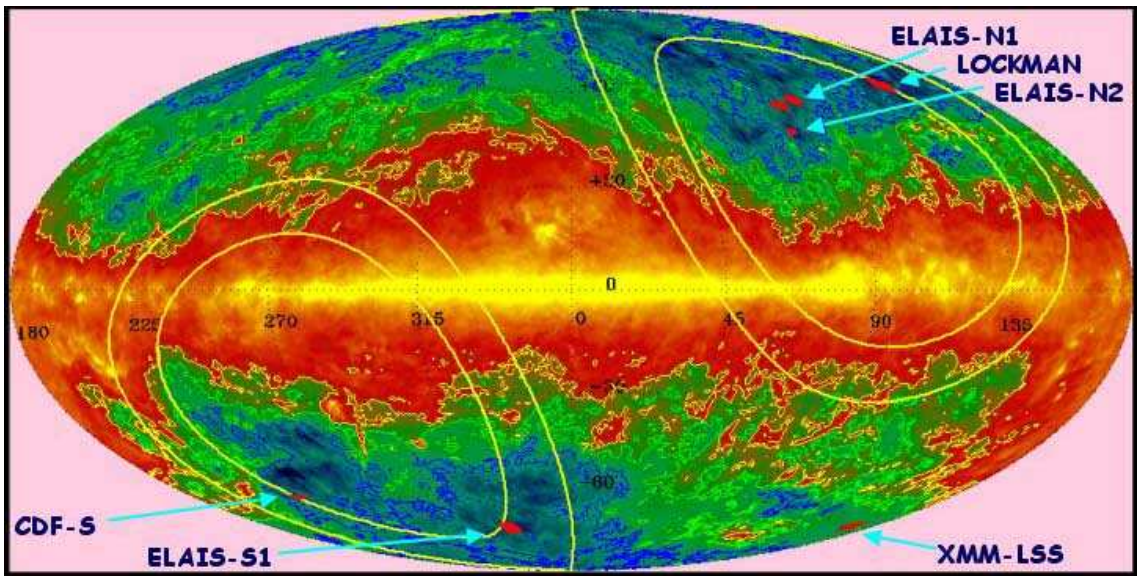


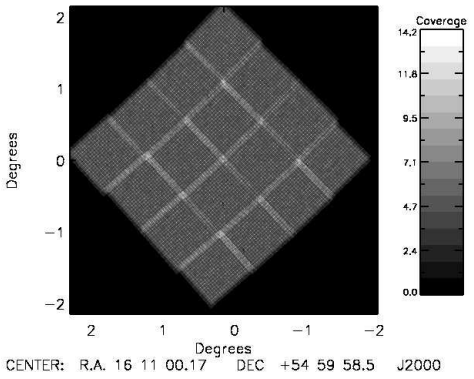
Figure 3.1: Hammer-Aitoff equal area projection in galactic coordinates of the sky imaged at 100μm, Schlegel, Finkbeiner & Davis (1998). The SWIRE fields are shown in red. Map by S. Oliver.

160μm coverage maps for five of the SWIRE fields.

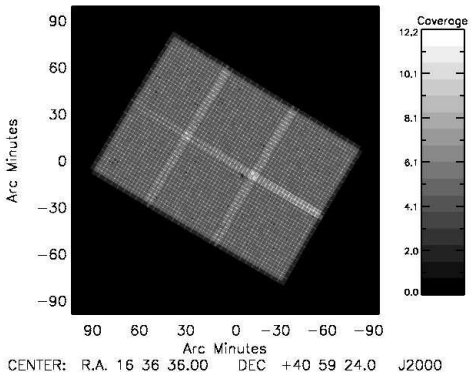
3.4.2 Completeness in SWIRE

As we move to fainter fluxes the fraction of sources detected decreases and it is important to know what that fraction is. This will vary depending on the observed waveband, flux and the coverage of the field which varies with position on the sky.

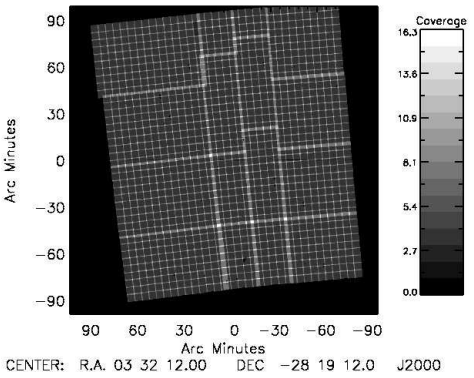
The completeness function tells us the fraction of sources detected in the survey as a function of flux. For an individual source the completeness can be thought of as the probability of detecting it. It is dependent on the source's location, as the completeness is a function of the noise in the image which varies with position, and flux. The coverage maps (Figures 3.2 and 3.3) were used to calculate completeness functions for ELAIS-N1 and ELAIS-N2 by Mattia Vaccari. The completeness function was calculated by simulating artificial sources and adding them into the SWIRE images. The source extraction stage of the analysis was then repeated, and the new list of resolved sources was compared with the known positions and fluxes of the artificial sources. The fraction of artificial sources that were recovered by the source extraction was computed as a function of flux and coverage (noise). Figure 3.6, shows the fraction of sources detected as a function of flux for the mean coverage across the field.



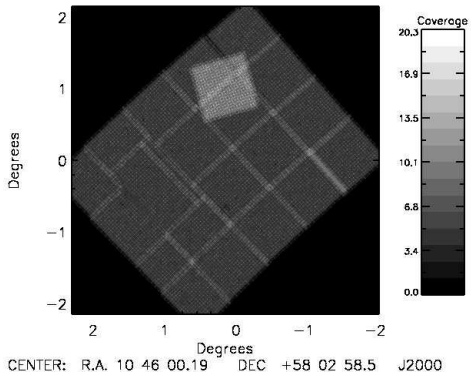
(a) ELAIS-N1



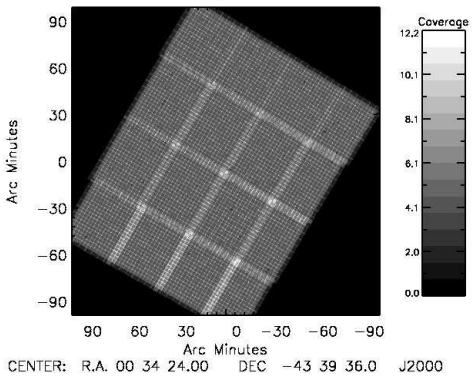
(b) ELAIS-N2



(c) CDFS

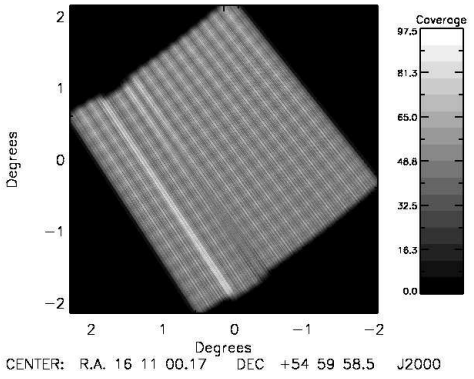


(d) Lockman hole

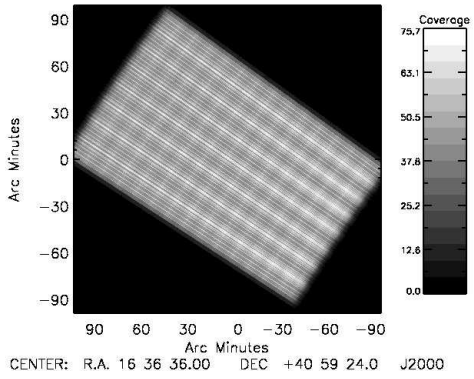


(e) ELAIS-S1

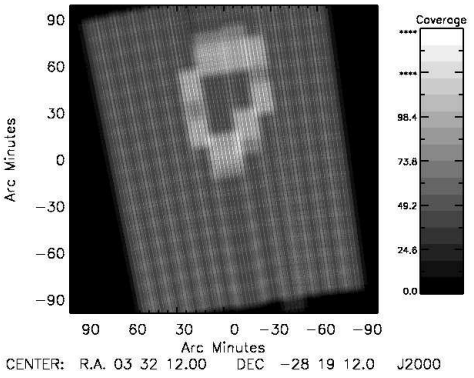
Figure 3.2: IRAC 3.6μm coverage maps for the five SWIRE fields ELAIS-N1, ELAIS-N2, ELAIS-S1, CDFS and Lockman hole. Black indicates no coverage and progressively lighter shades of grey indicate higher coverage, see scales adjacent to the maps for exact numbers.



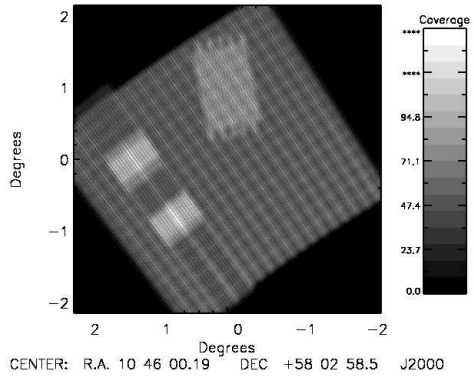
(a) ELAIS-N1



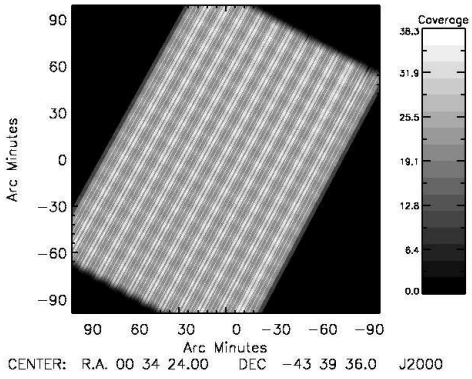
(b) ELAIS-N2



(c) CDFS

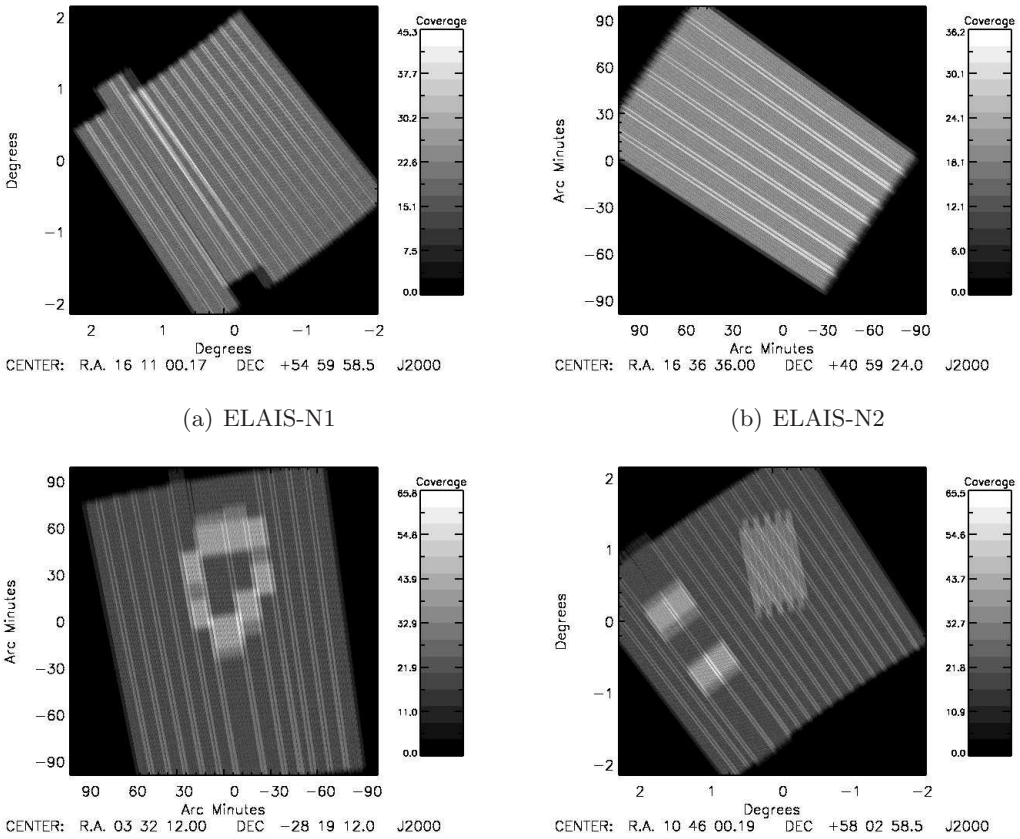


(d) Lockman hole



(e) ELAIS-S1

Figure 3.3: MIPS 24μm coverage maps for the five SWIRE fields ELAIS-N1, ELAIS-N2, ELAIS-S1, CDFS and Lockman hole. Black indicates no coverage and progressively lighter shades of grey indicate higher coverage (integration time).



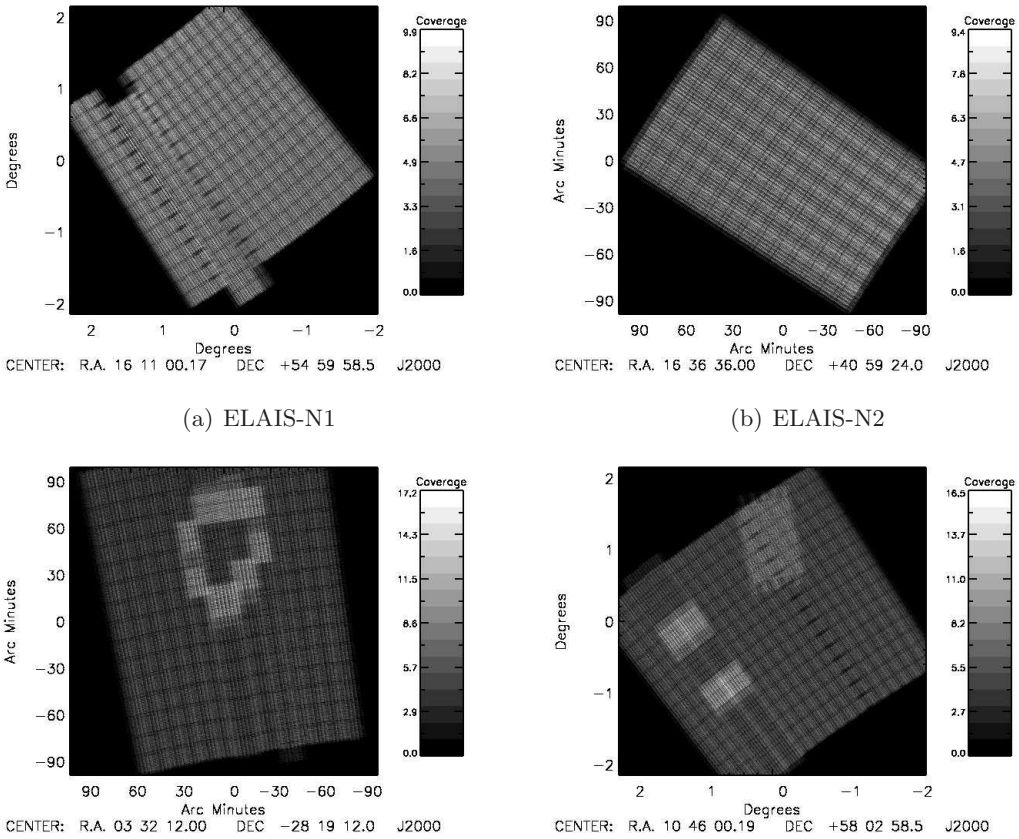
(a) ELAIS-N1

(b) ELAIS-N2

(c) CDFS

(d) Lockman hole

Figure 3.4: MIPS 70μm coverage maps for the five SWIRE fields ELAIS-N1, ELAIS-N2, ELAIS-S1, CDFS and Lockman hole. Black indicates no coverage and progressively lighter shades of grey indicate higher coverage (integration time).



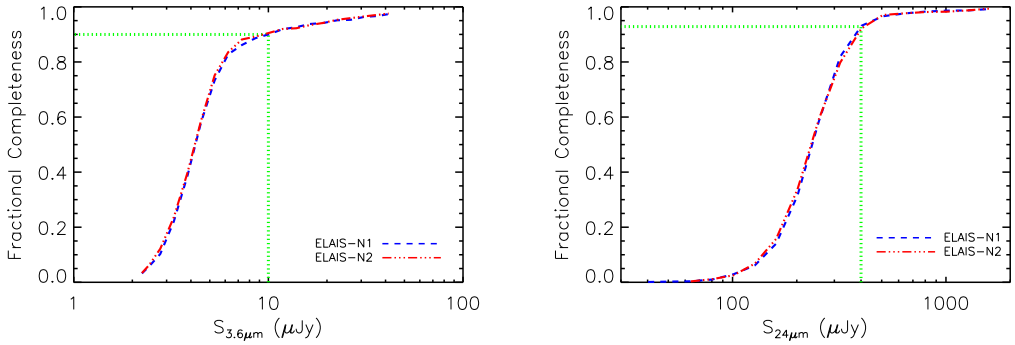
(a) ELAIS-N1

(b) ELAIS-N2

(c) CDFS

(d) Lockman hole

Figure 3.5: MIPS 160μm coverage maps for the five SWIRE fields ELAIS-N1, ELAIS-N2, ELAIS-S1, CDFS and Lockman hole. Black indicates no coverage and progressively lighter shades of grey indicate higher coverage (integration time).



(a) 3.6μm (b) 24μm

Figure 3.6: Completeness curves for ELAIS-N1 (blue dashed curve) and ELAIS-N2 (red dot-dashed curve) showing the fraction of sources detected as a function of a) $S_{3.6\mu m}$ and b) $S_{24\mu m}$ for the mean coverage (κ_{mean}=5.75) in the field. The green dotted lines indicate the completeness corresponding to the flux cuts applied to the catalogues used in the clustering analysis (see Section 3.7).

3.5 The SWIRE Photometric Redshift Catalogue

The SWIRE fields were chosen at locations on the sky where the most ancillary data was available. Partially coincident with ELAIS-N1 and ELAIS-N2 is the Isaac Newton Telescope Wide Field Survey (INTWFS), McMahon et al.(2001) and Lewis et al.(2000). The INT is a conventional Cassegrain telescope with a diameter of 2.54 meters located at La Palma. The WFS covers 100 deg^2 in up to six wavebands; u, B, g, r, i, z. See Table 5.1 for the area of each field with optical coverage.

Rowan-Robinson et al.(2008) analysed the SWIRE sources with optical associations from the WFS survey using the template fitting photometric redshift code IMPZ of Rowan-Robinson(2003) and subsequently Rowan-Robinson et al. 2005 and Babbedge et al.(2004) to produce the SWIRE photometric redshift catalogue containing 1,025,119 sources across all fields. The area of each field with available photometric redshifts was dictated by the coverage of optical data. A subset of 5976 galaxies in the photometric catalogue also had spectroscopic redshifts for which Rowan-Robinson et al.(2008) found a typical rms deviation to be $(z_{phot} - z_{spec})/(1 + z_{spec}) \sim 3.5\%$. Figure 3.7 shows the photometric redshifts plotted against the spectroscopic redshifts. The redshift resolution is 0.01 in $\log_{10}(1 + z)$ from 0-0.85 ($0 < z < 6$) providing 85 distinct redshift bins.

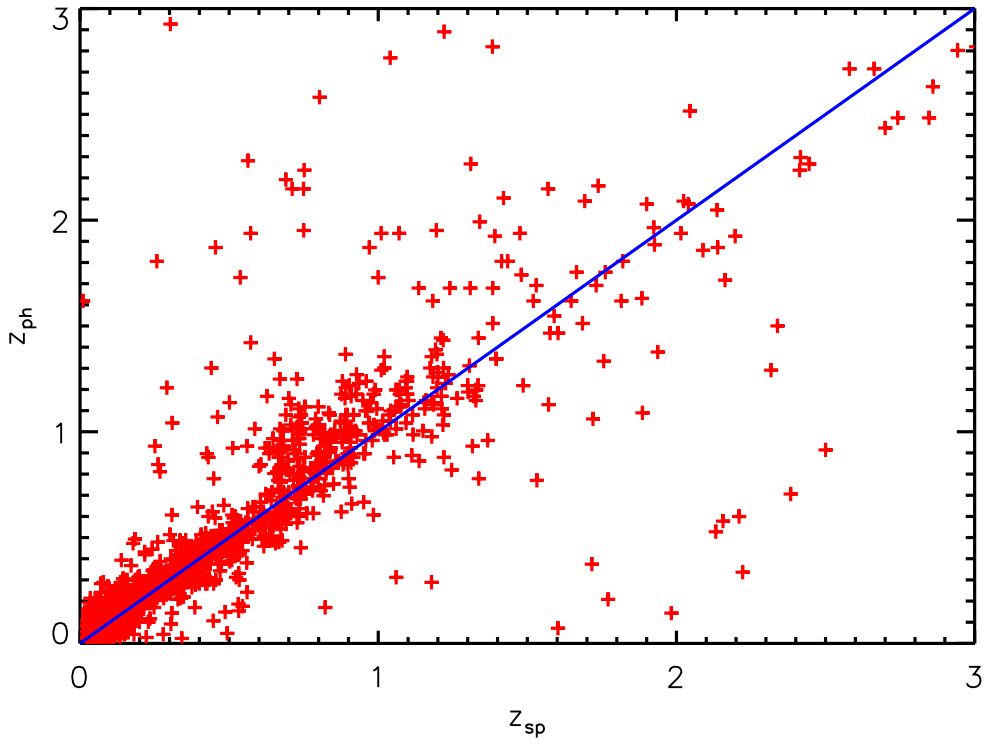


Figure 3.7: The photometric redshifts against the spectroscopic redshifts for galaxies with $r \leq 23.2$, $S_{3.6\mu m} \geq 10\mu$Jy, $n_{band} \geq 5$ and $\chi^2 \leq 5$. The blue line shows a 1:1 relation.

3.5.1 Template Fitting

Rowan-Robinson et al.(2008) started with empirical templates from Yoshii & Takahara(1988) for galaxies of type E, Sab, Sbc, Scd, Sdm and from Calzetti & Kinney(1992) for starburst galaxies, sb. These empirical templates were improved using the available spectroscopic data, for which many of the galaxies had 10-band photometry from the CFH12K-VIRMOS survey Le Fèvre et al.(2004). The latter were then regenerated to higher resolution using Simple Stellar Populations (SSPs), each weighted by a different SFR and extinguished by a different amount of dust, A_V. This procedure, based on the synthesis code of Poggianti et al.(2001), gave the templates a physical validity. Minimization was based on the Adaptive Simulated Annealing algorithm, the details of which along with the fitting technique are given in Berta et al.(2004).

Rowan-Robinson et al.(2008) also calculated additional properties for every galaxy in the photometric redshift catalogue including stellar mass and SFR estimates. For full details of the methods used we refer the reader to Rowan-Robinson et al.(2008). Below we summarize the key points in the derivation of these quantities.

3.5.2 Stellar Masses

For each galaxy Rowan-Robinson et al.(2008) estimated the rest-frame 3.6μm luminosity, $\nu L_\nu(3.6)$, in units of $\mathrm{L}_\odot$ and from their stellar synthesis models they estimated the ratio $(M_\star/\mathrm{M}_\odot)/(\nu\mathrm{L}_\nu(3.6)/\mathrm{L}_\odot)$, which they found to be 38.4, 40.8, 27.6, 35.3, 18.7 and 26.7 for galaxy types E, Sab, Sbc, Scd, Sdm and sb, respectively. Estimates based on the B-band luminosity gave values of $M_\star$ agreeing with these within 10-20%. Estimates based on 3.6μm luminosity should be more reliable, since there is a better sampling of lower mass stars and less susceptibility to recently formed massive stars. These mass estimates are strictly valid only for low redshift. For higher redshifts the mass-to-light estimates will be lower since for the oldest stellar populations, M/L varies strongly with age (Bruzual A. & Charlot(1993), see their Figure 3). This can be approximately modeled using the Berta et al.(2004) synthesis fits described above, with an accuracy of 10%, as

$$\frac{(M_\star/\mathrm{M}_\odot)}{(\nu L_\nu(3.6)/\mathrm{L}_\odot)(\mathrm{t})} = \frac{50}{(a + 1.17(t/t_0)^{-0.6})} \tag{3.3}$$

where t_0 is the present epoch and a = 0.15, 0.08, 0.61, 0.26, 1.44, 0.70 for SED types E, Sab, Sbc, Scd, Sdm and sb, respectively. This approach should correctly capture the

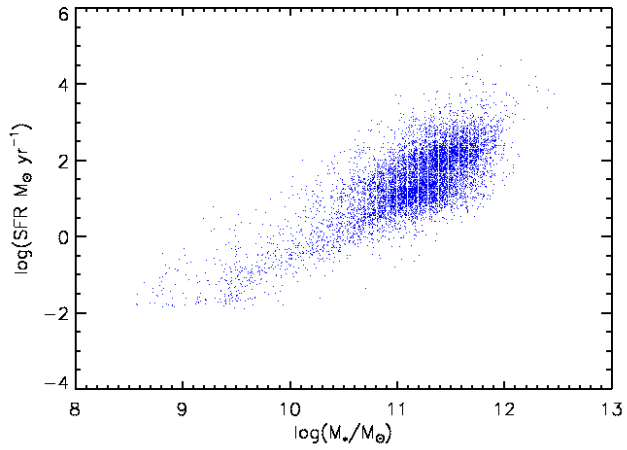


Figure 3.8: SFR against stellar mass for spiral galaxies in the photometric redshift catalogue.

different evolutionary behaviour of stellar masses in star-forming galaxies and of early-type galaxies.

3.5.3 Star Formation Rates

The photometric catalogue also contained star formation rates for sources with 24μm detections. The SFR was estimated using the conversion from 60μm luminosity (which was derived from the template fits) of Rowan-Robinson et al.(1997), Rowan-Robinson(2001),

$$SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1} = 2.2\epsilon^{-1}10^{-10}\frac{\mathrm{L}_{60}}{\mathrm{L}_{\odot}} \tag{3.4}$$

where ϵ is the fraction of UV light absorbed by dust and taken to be 2/3 [note the discussion in Rowan-Robinson(2003) and Bell(2003) whereby illumination of ambient dust by older stars leads to increased far infrared emission which can lead to an overestimation of the star-formation rate]. The bolometric corrections at 60μm needed to convert L_{ir} to L_{60}, are 3.48, 1.67 and 1.43 for cirrus, M82 and A220 templates, respectively. Figure 3.8 shows the relationship between SFR and stellar mass.

3.5.4 Bayesian Prior for the Redshift Distribution

It is important to use the best possible estimate of the redshift distribution when using Limber's equation (Equation 2.21) to calculate the spatial clustering strength r_0 (see Section 2.5). The following section outlines how an improved estimate of the redshift

distribution can be made over the simple $N(z)$ (i.e. just using the photometric redshifts as they are in the catalogue).

The photometric redshift distribution is spread across 85 discrete redshifts equally spaced by 0.01 in $\log(1+z)$ from 0.01 to 0.85. Rowan-Robinson et al.(2008) calculated a χ^2 to quantify how good the template fit was at each redshift. An estimate of the redshift distribution was made by converting the full reduced χ^2 distribution for galaxy i into a probability density function (PDF) using,

$$PDF_i = \frac{(1/2)^{\nu_i/2}}{\Gamma(\nu_i/2)} \; (\chi_i^2)^{\nu_i/2-1} \; e^{(-\chi_i^2/2)} \tag{3.5}$$

where ν was the number of degrees of freedom in the template fit and Γ was the standard gamma function. The PDF was then summed over all galaxies to give an estimate of the redshift distribution,

$$N_p(z) = \sum_i^{N_{gal}} PDF_i \tag{3.6}$$

$N_p(z)$ is shown as the dotted line in Figure 3.9. By using the probability distribution, $N_p(z)$ improved on the simple $N(z)$ by taking into account that some galaxies might have had a good probability of being detected at more than one redshift, e.g. $p(z=0.1)=0.48$ and $p(z=0.6)=0.52$, i.e. it takes into account the redshift uncertainties.

The PDF$_i$ gives the probability that the SED template, T, was correct given the photometric fluxes, S, i.e. $p(T \mid S)$. More specifically what we require is the probability that the template and redshift, z, was correct for galaxy i given the flux data, i.e. $p(z, T \mid S)$. This is related to $p(T \mid S)$ through the product rule which takes into account prior information $p(z \mid T, S)$,

$$p(z, T \mid S) = p(T \mid S) \; p(z \mid T, S) \tag{3.7}$$

Therefore to determine $p(z, T \mid S)$ we need to calculate the *prior* $p(z \mid T, S)$. We calculate this as follows. The probability of finding a galaxy of flux $S_{\nu 0}$ at redshift z in a redshift range dz fitted with a template T is,

$$p(z \mid T, S_{\nu 0}) \; dz = \frac{dN}{\int dN} \tag{3.8}$$

i.e. the probability of finding dN galaxies in redshift range dz is the fraction of the total number of galaxies N from $0 < z < \infty$ in the limit $dz \to 0$. The number of galaxies dN expected in volume element dV with luminosity range $d\log(L)$ is given by

$$dN = \phi(L_{(1+z)\nu 0}, T, z) \; dV \; d\log(L) \tag{3.9}$$

where $\phi(L_{(1+z)\nu 0}, T, z)$ is the luminosity function for galaxies with luminosity $L_{(1+z)\nu 0}$ fitted with template T at redshift z. For a survey of solid angle $d\Omega$ in redshift range dz this can be written as

$$dN = \phi(L_{(1+z)\nu 0}, T, z)\ \frac{dV}{d\Omega dz} d\Omega\ dz\ d\log(L) \tag{3.10}$$

Substituting Equation 3.10 into 3.8 gives

$$p(z \mid T, S_{\nu 0})\ dz = \frac{\phi(L_{(1+z)\nu 0}, T, z)\frac{dV}{d\Omega dz}\ d\Omega\ dz\ d\log(L)}{\int \phi(L_{(1+z)\nu 0}, T, z)\frac{dV}{d\Omega dz}\ d\Omega\ dz\ d\log(L)} \tag{3.11}$$

where $d\Omega$ and $d\log(L)$ do not vary with z, hence

$$p(z \mid T, S_{\nu 0}) = \frac{\phi(L_{(1+z)\nu 0}, T, z)\frac{dV}{d\Omega dz}}{\int \phi(L_{(1+z)\nu 0}, T, z)\frac{dV}{d\Omega dz} dz} \tag{3.12}$$

To evaluate Equation 3.12 we used the 3.6μm and 24μm luminosity functions determined by Babbedge et al.(2006) where the evolution with redshift was also parameterized. Babbedge et al.(2006) modelled the 3.6μm luminosity function as a Schechter function

$$\phi(L)\frac{dL}{L^\star} = \phi^\star \left(\frac{L}{L^\star}\right)^\alpha e^{-L/L^\star} \frac{dL}{L^\star} \tag{3.13}$$

and the 24μm luminosity functions as a double power law

$$\phi(L)\frac{dL}{L^\star} = \phi^\star \left(\frac{L}{L^\star}\right)^{1-\alpha} \frac{dL}{L^\star}\ ,\ L < L^\star \tag{3.14}$$

and

$$\phi(L)\frac{dL}{L^\star} = \phi^\star \left(\frac{L}{L^\star}\right)^{1-\beta} \frac{dL}{L^\star}\ ,\ L > L^\star \tag{3.15}$$

The values for the parameters for $\phi^\star$, $L^\star$, α and β are given in Table 3.2

We calculated the prior using Equation 3.12, which we then substituted into Equation 3.7 to determine $p(z, T \mid S)$. The top panel of Figure 3.9 shows the redshift distribution with and without the prior for all types of galaxies combined. The lower panel shows the redshift distribution for elliptical, spiral and starburst galaxies separately. ELAIS-N1 and ELAIS-N2 both show a clear peak at $z \sim 0.5$. The prior acts to narrow this peak slightly for spiral galaxies and broadens it very slightly for starburst galaxies. At higher z the number of objects best fit by elliptical templates declines dramatically while the spiral and starburst templates increase with a clear second peak at $z \sim 0.9$. The prior acts to increase the height of this second peak for starburst galaxies. This second peak can be explained by considering the effect of the $r \leq 23.2$ constraint (see Section 3.7.3)

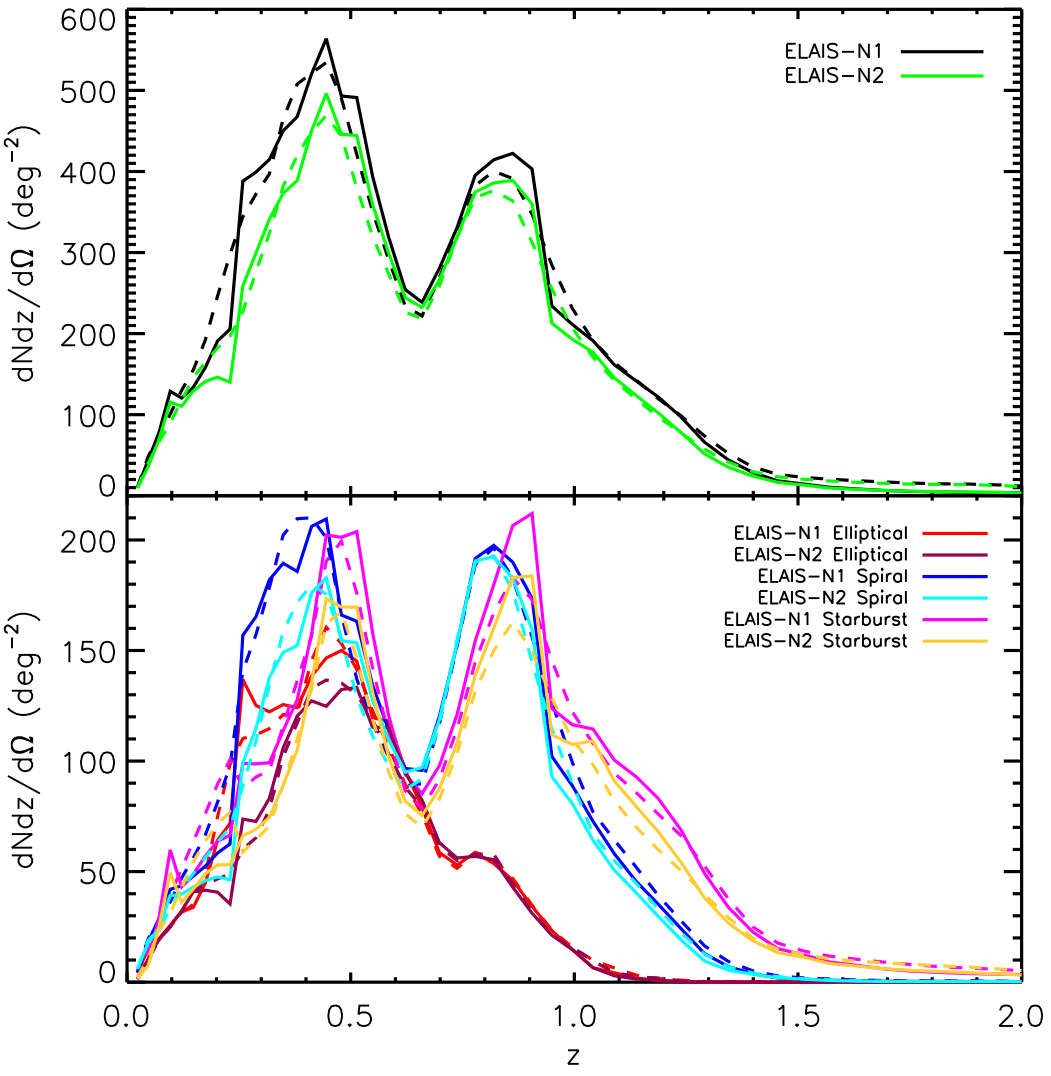


Figure 3.9: The redshift distribution of galaxies in the photometric redshift catalogue for ELAIS-N1 and ELAIS-N2. The top panel shows the number of galaxies per redshift interval per square degree without the prior (dashed line) and with the prior included (solid line). The lower panel shows the same except the galaxies are broken down into ellipticals (template E), spirals (templates Sab-Scd) and starburst (templates Sb-Sdm) galaxies.

Table 3.2: Schechter function and power law parameters for SWIRE luminosity functions determined by Babbedge et al.(2006). 3.6μm (Schechter function) and 24μm (double power law) luminosity function parameters are given for 4 different redshift ranges.

Sample	α	β	$L^\star$	$\phi^\star$	Redshift
			$h^{-2}L_\odot$	h^3 Mpc^{-3} dex^{-1}	
3.6μm	-0.9	-	$5.6^{+7.1}_{-4.5} \times 10^9$	3.6×10^{-3}	0-0.25
3.6μm	-1.0	-	$7.9^{+8.9}_{-7.1} \times 10^9$	4.5×10^{-3}	0.25-0.5
3.6μm	-0.9	-	$6.8^{+9.1}_{-7.9} \times 10^9$	7.0×10^{-3}	0.5-1.0
3.6μm	-0.9	-	$7.1^{+7.9}_{-5.6} \times 10^9$	4.0×10^{-3}	1.0-1.5
24μm	1.3	2.5	$3.6^{+6.3}_{-1.8} \times 10^8$	5.0×10^{-2}	0-0.25
24μm	1.3	3.0	$1.6^{+1.9}_{-1.3} \times 10^9$	5.0×10^{-2}	0.25-0.5
24μm	1.3	3.0	$3.2^{+3.6}_{-1.1} \times 10^9$	5.0×10^{-2}	0.5-1.0
24μm	1.3	3.0	$4.5^{+5.6}_{-3.0} \times 10^9$	5.0×10^{-2}	1.0-1.5

as the different SEDs are redshifted through the observed filters. At $z \sim 0.7$ the strong 4000Å break feature moves from the r to the i band, dramatically decreasing the flux in the r band. This means that only the most luminous galaxies remain in the sample after imposing the $r \leq 23.2$ constraint. The second peak of spiral types occurs because at $z \sim 0.9$ the observed frame g maps onto the strong far UV (< 2000Å) flux associated with ongoing star formation. Early-type galaxies typically do not have strong ongoing star formation and so do not exhibit this second peak.

The redshift distribution (dN/dz) for each ($M_\star, z$) or (SFR, z) cell used in Equation 2.21 to calculate the spatial clustering strength r_0 were constructed using the improved estimate of the redshift distribution including the prior.

3.6 Optical Completeness Curves

The optical data from the INTWFS, specifically the u, g, r, i, z bands, is of variable "depth" and completeness across each SWIRE field. The variable depth of the data arises from the observations being taken in a series of tiles of varying integration times and observing conditions, see Figures 3.10 and 3.11. The *tiles* in these figures show the 5σ magnitude limits for that part of the sky, i.e. sources brighter than the magnitude limit are reliable detections which were calculated by our collaborator Eduardo Gonzalez Solares. The

completeness for each optical band as a function of position is unknown. For a clustering analysis it is important to know what the completeness is so that variations in completeness can be minimised to prevent artificial clumpyness. It is particularly important to determine the r band completeness as galaxy samples in the clustering analysis are selected based on their r-band magnitude (see Section 3.7.3).

The completeness can be estimated by considering the number density of galaxies as follows. For a sample of galaxies i of magnitude m in the survey, the observed number density, $n_i(m)'$ is,

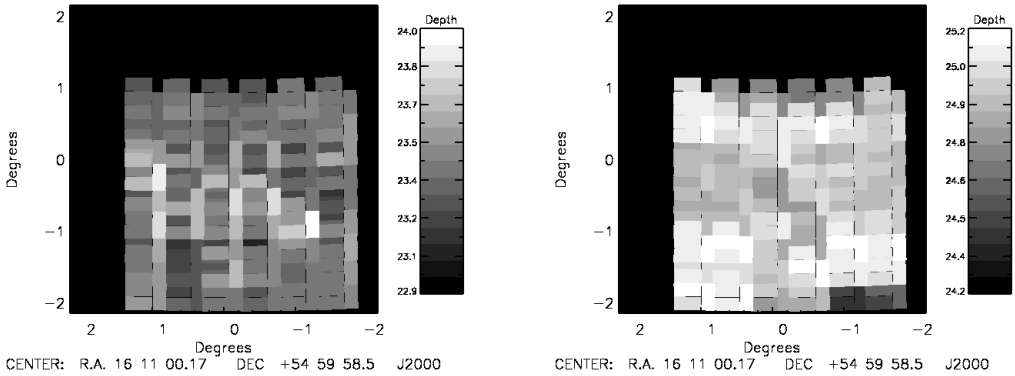
$$n_i(m)'(\alpha, \delta) = p_i(m, \alpha, \delta)\ n_i(m) \tag{3.16}$$

where $p_i(m)$ is the probability of detecting galaxies i (due to the selection effects of the survey) which varies with position (α, δ) due to the varying depth from tile to tile. $n_i(m)$ is the *true* number density of galaxies of magnitude m. $p_i(m)$ can be thought of the completeness of our sample.

Before the completeness could be estimated using this approach, we split each field into *regions*. For each waveband, we grouped the tiles together into 6 *regions* of approximately equal depth. This ensured there was a sufficient number of galaxies in each region to reliably measure the number densities. The regions were not necessarily the same for each waveband. We calculated the area of each region by counting the number of pixels on the tile map in each region and multiplying by the area of a pixel ($3'' \times 3''$). We estimated the completeness for each waveband and region as follows.

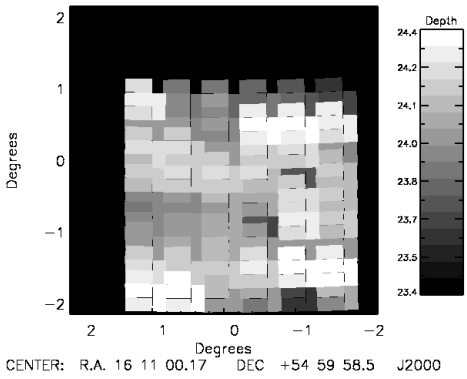
3.6.1 Estimating the Optical Completeness

Two data sets were used. The full optical catalogues from the INTWFS survey, which we will refer to as the *shallow* catalogues, and an approximately 1 magnitude deeper INT catalogue, which we will refer to as the *deep* catalogue. The deep catalogue was sufficiently deep to be considered to be complete to $r \sim 24$. Therefore it was used to represent the true number counts. The shallow catalogues were the optical catalogues which Rowan-Robinson et al.(2008) cross matched with the SWIRE catalogues before doing his template fitting procedure. The deep catalogue was coincident with part of the shallow catalogue and covered 0.25 deg^2. We removed the stars from both the shallow and deep catalogues following the method outlined in Section 3.7.2. We calculated the differential number

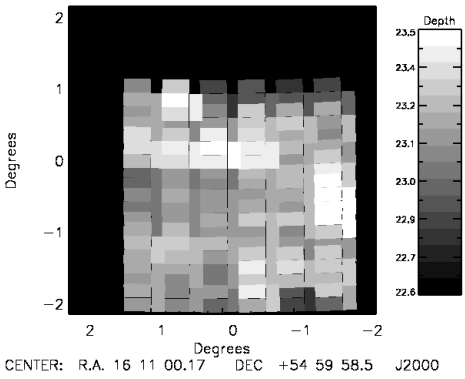


(a) u

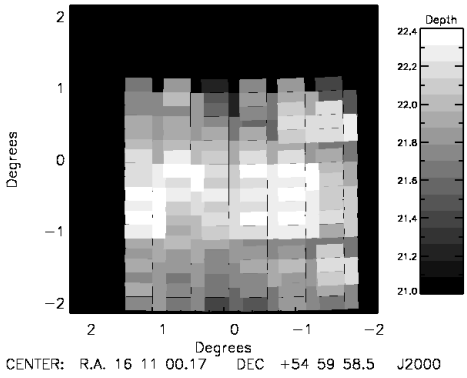
(b) g



(c) r

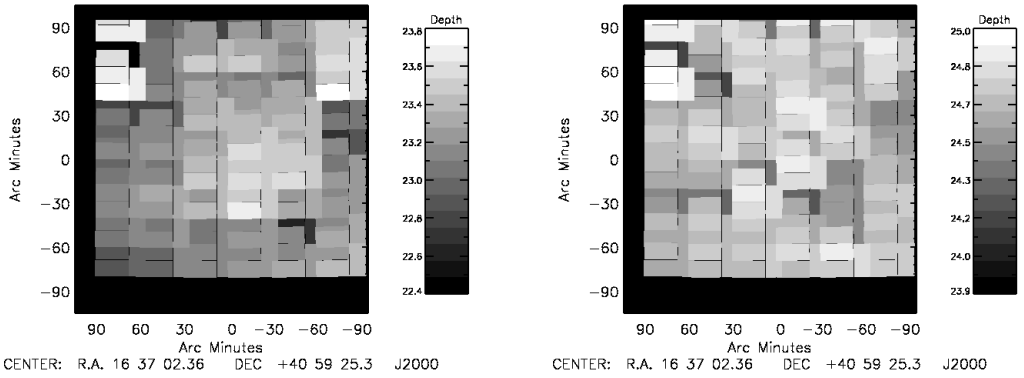


(d) i



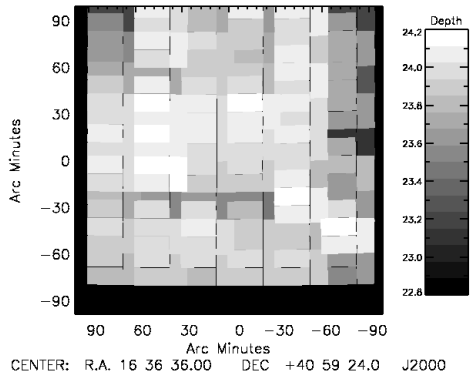
(e) z

Figure 3.10: INTWFS u, g, r, i, z-band tile maps for ELAIS-N1. Black indicates no coverage and progressively lighter shades of grey indicate deeper parts of the field.

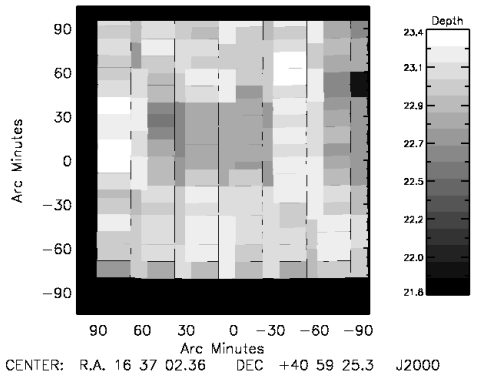


(a) u

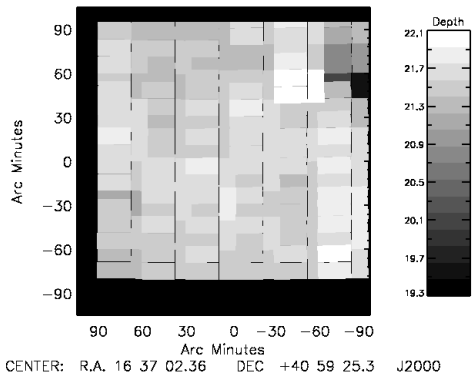
(b) g



(c) r



(d) i



(e) z

Figure 3.11: INTWFS u, g, r, i, z-band tile maps for ELAIS-N2. Black indicates no coverage and progressively lighter shades of grey indicate deeper parts of the field.

Waveband	R1	R2	R3	R4	R5	R6
g	0.07	0.40	1.20	1.56	1.58	0.75
r	0.12	0.15	1.15	1.23	2.23	0.69
i	0.36	1.38	1.84	1.12	0.85	-

Table 3.3: The area of each region in square degrees in ELAIS-N1 for each waveband used to make the optical completeness curves.

density of galaxies in the deep catalogue as a function of magnitude using,

$$N_d(m) = \frac{n_d(m)}{d\Omega_d} \tag{3.17}$$

where $n_d(m)$ was the number of galaxies of magnitude m and $d\Omega_d$ was the area of the deep catalogue. Then for region R (for a given waveband) the number density of galaxies in the shallow catalogue was calculated using,

$$N_R(m) = \frac{n_R(m)}{d\Omega_R} \tag{3.18}$$

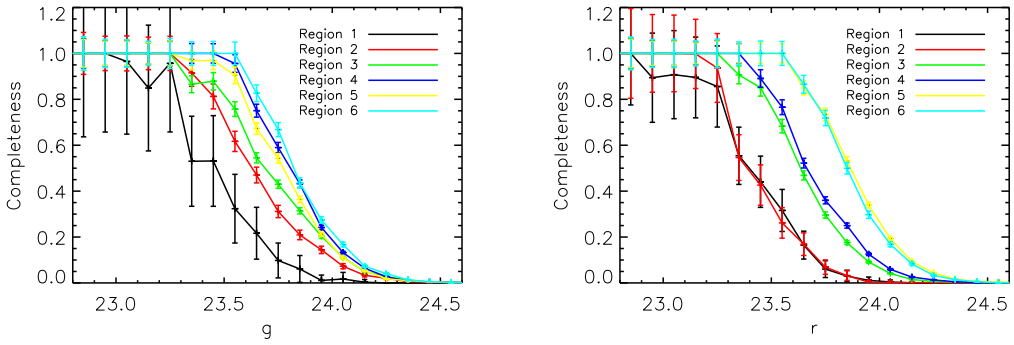
where $n_R(m)$ was the number of galaxies of magnitude m in region R and $d\Omega_R$ is the area of region R. The completeness in region R was then given by

$$C = \frac{N_R(S)}{N_d(S)} \tag{3.19}$$

We repeated this for each region to produce a completeness curve for each region (for each waveband). Figure 3.12 shows the optical completeness curves for g, r, i wavebands for each region in ELAIS-N1. The completeness curves for ELAIS-N2 look very similar. Region 1 was the shallowest region and region 6 was the deepest. Region 1 for the g band covered a small area of 0.07 deg^2 leading to larger errors on the number density and hence larger errors on the completeness. Only 5 regions were used for the i band to provide more galaxies in each region. The r band was 100% complete by $r = 23.3$ in regions 3-6. Regions 1 and 2 were 90% complete by $r = 23.2$. These regions covered a relatively small area of 0.12 and 0.15 deg^2, respectively, see Table 3.3.

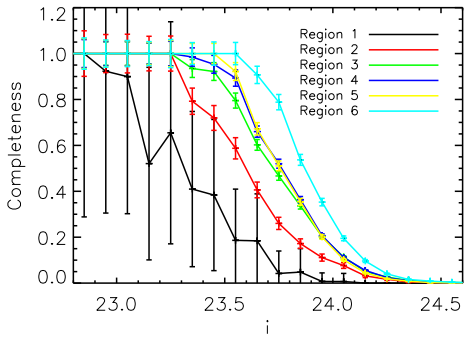
3.7 Selection Effects in the Photometric Redshift Catalogue

The clustering analysis described in detail in Chapter 5, requires that the selection effects inherent to the catalogue are satisfactorily understood. The photometric redshift catalogue



(a) g

(b) r



(c) i

Figure 3.12: Optical completeness curves for wavebands g, r, i. Each region is plotted as a different colour indicated by the legend in each plot. Region 1 is the shallowest region and region 6 is the deepest. Only 5 regions were used for the i band. Deeper regions are more complete at fainter magnitudes as expected.

combines data from SWIRE and the INTWFS surveys and hence it's selection effects are a combination of the individual survey's selection effects. Variations in completeness in either SWIRE at 3.6μm *or* 24μm or INTWFS r-band (which is a fairy good tracer of photo-z accuracy, see Figure 10 of Rowan-Robinson et al.(2008)) could produce an artificial clustering signal which we would not be able to determine from the real signal. The following sections outline how we defined criteria to constrain the selection effects due to variations in completeness in SWIRE, variations in r-band completeness and to select only "good" photometric redshifts. To fully appreciate the following discussion we refer the reader to Section 5.2 for a description of the $(M_\star, z)$ and (SFR, z) analyses.

3.7.1 Minimizing the Variation in SWIRE Completeness

Locations on the sky with a higher completeness contained more sources and hence appeared more clustered simply because the field was deeper at those points. Therefore, we minimized the variation in completeness (as a function of position) to prevent introducing an artificial clustering signal. Equation We know the completeness is a function of coverage (which is position dependent) and flux. We calculated the completeness for every point (pixel) on the sky as a function of flux. We explored the combined effects of various coverage thresholds, κ_T, and flux limits S_c on the mean completeness and the variation in completeness. Limits were chosen which ensured the average completeness across the field was high and the variation in completeness was small to prevent artificial clumpyness.

For the $M_\star - z$ analysis a 3.6μm flux cut of 10μJy and a coverage threshold of 4 were applied which ensured both fields were $> 90\%$ complete and the variation in completeness was $< 2\%$, see Figure 3.6(a) and Table 3.4. For the $SFR - z$ analysis a $S_{24\mu\mathrm{m}}$ cut of 400μJy and a coverage threshold of 30 were applied making ELAIS-N1 and ELAIS-N2 93% and 92% complete, respectively, see Figure 3.6(b), with the variation in completeness $< 4\%$, see Table 3.5. We applied the flux cuts directly to the catalogues and the coverage thresholds were applied by making a binary mask for each field and waveband.

Each mask measured 5160×5160 pixels where each pixel measured $3'' \times 3''$ and every pixel initially had a value of 1. We defined coordinates for the corners of the fields by eye and joined them by straight lines to define the boundaries of the mask. All pixels outside the boundaries of the field were set to 0. The coverage maps were used to identify pixels with a coverage value below the coverage threshold and such pixels were masked out by being set to 0. In total four masks were constructed; 3.6μm and 24μm for ELAIS-N1 and

Field	$S_c(\mu\text{Jy})$	κ_T	$\bar{C}(\%)$	σ_C (%)
ELAIS-N1	10	4.0	90	1.2
ELAIS-N2	10	4.0	90	1.0

Table 3.4: The mean completeness and variation in completeness in ELAIS-N1 and ELAIS-N2 resulting from the $S_{3.6\mu m}$ and coverage cuts. The flux cuts (S_c) and coverage thresholds (κ_T) were defined to give a mean completeness, $\bar{C}$ and a small variation in completeness, σ_C, at 3.6μm.

Field	$S_c(\mu\text{Jy})$	κ_T	$\bar{C}$ (%)	σ_C (%)
ELAIS-N1	400	30	93	3.0
ELAIS-N2	400	30	92	3.5

Table 3.5: The mean completeness and variation in completeness in ELAIS-N1 and ELAIS-N2 resulting from the $S_{24\mu m}$ and coverage cuts. The flux cuts (S_c) and coverage thresholds (κ_T) were defined to give a mean completeness, $\bar{C}$ and a small variation in completeness, σ_C, at 24μm.

ELAIS-N2, see Figure 3.13.

3.7.2 Removing Foreground Stars

Foreground stars typically act to dilute the clustering signal as they obscure galaxies along their line-of-sight, hence precluding them from the analysis. We identified foreground stars following the method of Waddington et al.(2007). We cross-matched the SWIRE catalogues with the 2MASS point source catalogue (Skrutskie et al.(2006)) to classify the bright sources. Any bright point source with $K < 12$ was identified to be a star. The faint 2MASS sources were classified based on their near-to mid infrared colours and stellarity at 3.6μm. Point-like (STELLARITY$>$ 0.94) sources were identified as stars. Sources that were both blue ($J - m_{36} < -1.50$ or $H - m_{36} < -2.2$) and not clearly resolved were also stars. Stars fainter than 2MASS were classified based on their mid-infrared colours and 3.6μm stellarity, where blue compact sources were stars. A circular region of radius $\log_{10}(R) = 3.1 - 0.16K$ arcsec was measured around each star. We determined this,

by visual inspection, to be the point at which the stars point source function became indistinguishable from the background. All the circular regions were transformed into pixel coordinates and incorporated into the binary masks by setting all pixels which fell within each circle to 0, hence excluding the stars and nearby sources from the analysis.

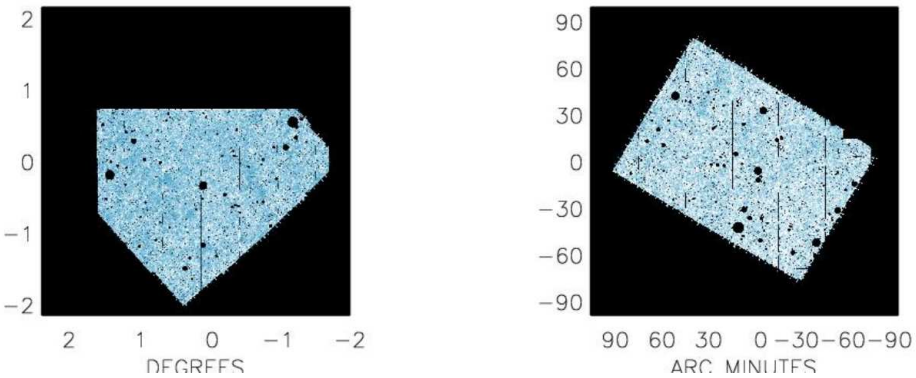
3.7.3 Minimizing the Variation in r-band Completeness

The fraction of photometric redshift outliers increases at faint r-band magnitudes, Rowan-Robinson et al.(2008). Therefore we imposed a magnitude limit of $r_{cut} = 23.2$ to improve the photometric redshift accuracy. The r band completeness curve (Figure 3.12(b)) shows at $r \leq 23.2$ the catalogue is $\sim 90\%$ in shallow areas of the field (which comprise a small area) and approaching 100% complete in other areas. We made an r-band optical binary mask based on the tile maps which showed the varying depth of the r-band data across the field. Shallow tiles were excluded from the analysis. We set all pixels within such tiles to 0 and all other pixels to 1. After these areas were removed, the variation in completeness was $\sim 1\%$. The optical mask was then combined with the binary mask.

Figure 3.13 shows the binary masks with all components; the SWIRE coverage cut and optical r band depth cuts applied and the foreground stars removed. Strictly speaking these are 3.6μm masks however the 24μm look similar. The masks enabled position dependent criteria to be easily applied to the catalogues to produce homogeneous data sets which is critical for a clustering analysis.

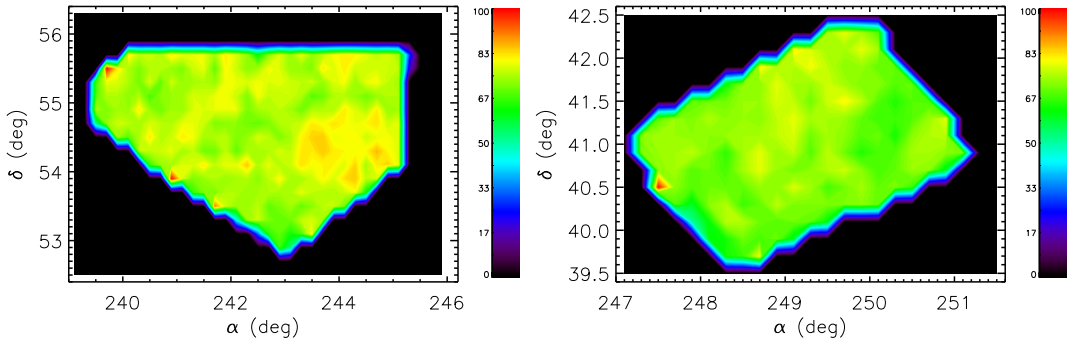
3.7.4 Minimising Photometric Redshift Outliers

The SED template fitting method of Rowan-Robinson et al.(2008) utilized multiple optical bands, the coverage of which was not homogeneous across each field. Therefore the number of photometric bands, n_{band}, in which each galaxy was detected varied. The goodness of fit of each SED template was quantified by a χ^2. Rowan-Robinson et al.(2008) recommended $\chi^2 \leq 5$ and $n_{band} \geq 5$ for accurate photometric redshifts. Imposing these constraints on the catalogues resulted in a large artificial variation in the number density of sources across each field which would have affected the clustering signal. Figures 3.14 show the percentage of galaxies as a function of position across the field satisfying these constraints. The strength of the variation produced an artificial clustering signal which would be reflected in the correlation function. Naively we would expect this affect to be less severe for galaxies with large fluxes as they are more likely to be detected in many wavebands and hence have



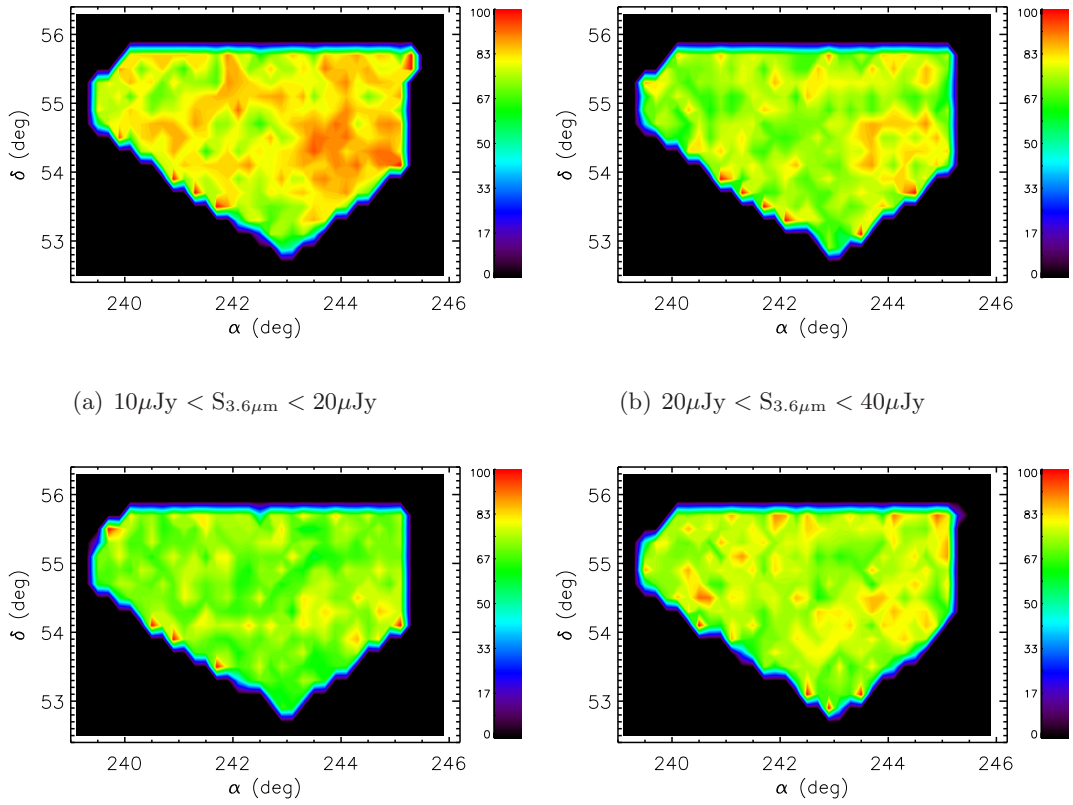
(a) ELAIS-N1 (b) ELAIS-N2

Figure 3.13: Binary masks for the SWIRE fields ELAIS-N1 and ELAIS-N2. Regions of the field included in the analysis are coloured white, representing a 1. Regions excluded from the analysis are coloured black, representing a 0. The light blue dots are the SWIRE galaxies overplotted on the masks.



(a) ELAIS-N1 (b) ELAIS-N2

Figure 3.14: Percentage of galaxies as a function of position satisfying the photometric constraints: $\chi^2 \leq 5$ and $n_{band} \geq 5$. The blue boundary is an artifact of the plotting.



(a) $10\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 20\mu\text{Jy}$

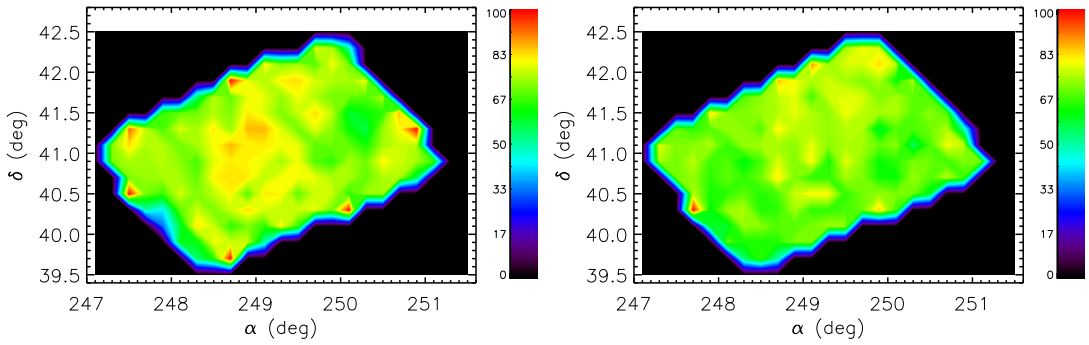
(b) $20\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 40\mu\text{Jy}$

(c) $40\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 80\mu\text{Jy}$

(d) $80\mu\text{Jy} < \text{S}_{3.6\mu\text{m}}$

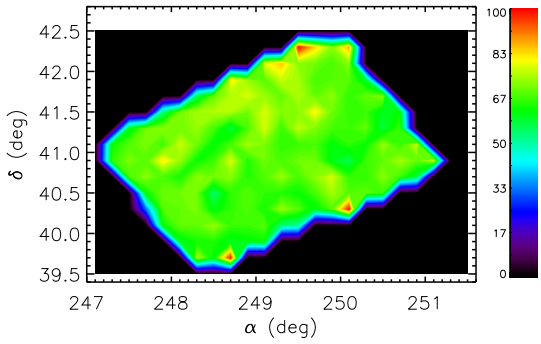
Figure 3.15: Percentage of galaxies as a function position in ELAIS-N1 with $\chi^2 \leq 5$ and $n_{band} \geq 5$. Panels a) to d) show the effect for galaxies in progressively brighter flux bins. The blue boundary is an artifact of the plotting.

a better template fit and therefore a more reliable photometric redshift, however this is found not to be the case. Figures 3.15 and 3.16 show the fraction of galaxies with $n_{band} \geq 5$ and $\chi^2 \leq 5$ in flux bins of increasing brightness for ELAIS-N1 and ELAIS-N2, respectively. These figures show there is no net improvement in the fraction of galaxies satisfying the photometric constraints $\chi^2 \leq 5$ and $n_{band} \geq 5$ with increasing flux. This suggests that some galaxies with large n_{band} expose the naivety of the SED templates and hence have a poor χ^2 fit to the template. Figures 3.17 and 3.18 show the fraction of galaxies satisfying just the $\chi^2 \leq 5$ constraint in increasing flux bins. There is no net change in the number of galaxies satisfying the χ^2 constraint for the 3 brightest flux bins, which suggests the brighter galaxies with larger n_{band} do not actually expose the naivety of the templates. However, the faintest flux bin shows a higher percentage of galaxies, on average across the

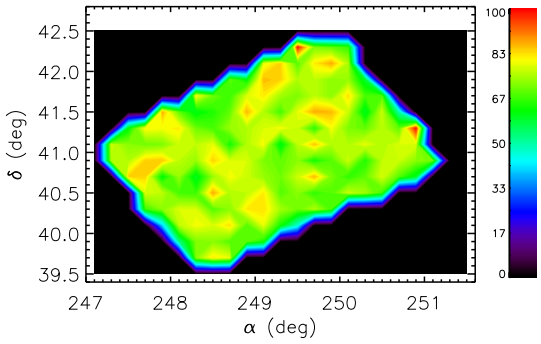


(a) $10\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 20\mu\text{Jy}$

(b) $20\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 40\mu\text{Jy}$

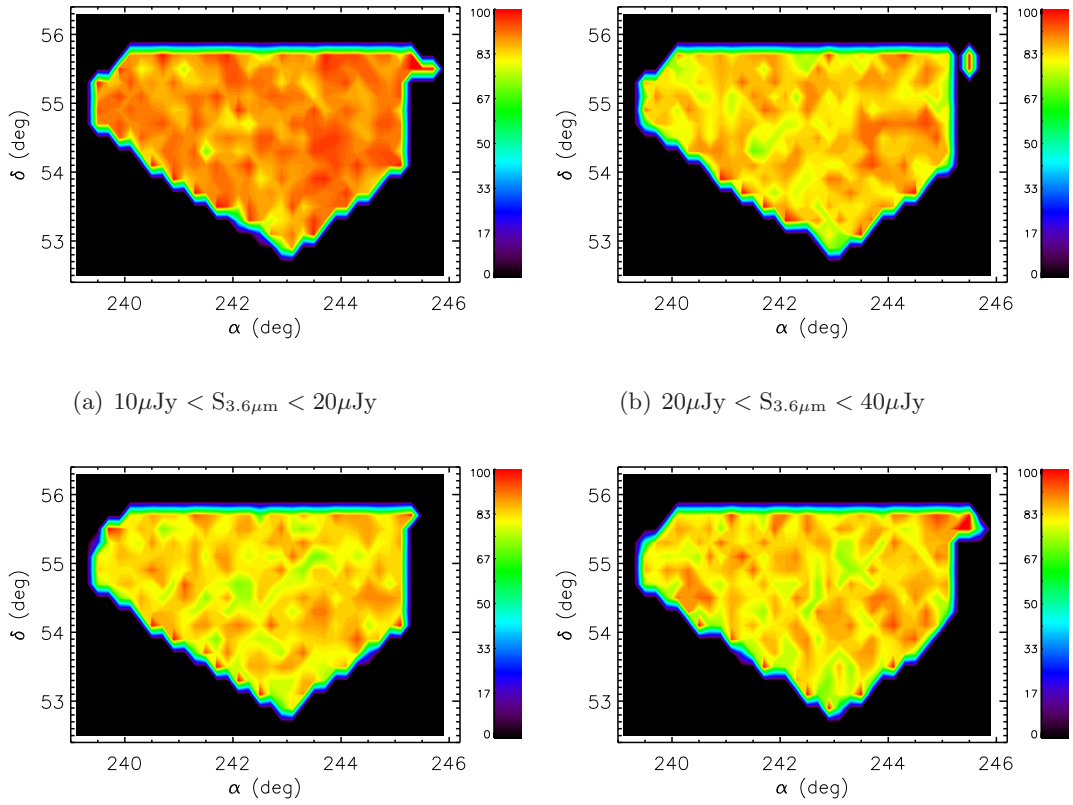


(c) $40\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 80\mu\text{Jy}$



(d) $80\mu\text{Jy} < \text{S}_{3.6\mu\text{m}}$

Figure 3.16: Percentage of galaxies as a function of position in ELAIS-N2 with $\chi^2 \leq 5$ and $n_{band} \geq 5$. Panels a) to d) show the effect for galaxies in progressively brighter flux bins. The blue boundary is an artifact of the plotting.



(a) $10\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 20\mu\text{Jy}$

(b) $20\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 40\mu\text{Jy}$

(c) $40\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 80\mu\text{Jy}$

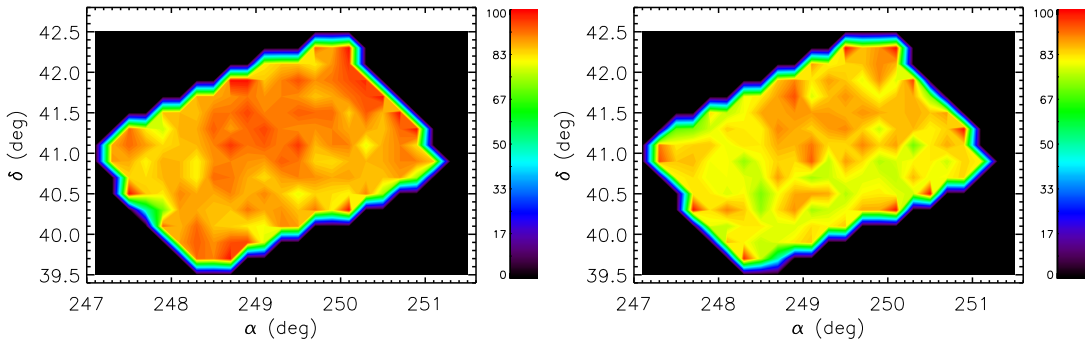
(d) $80\mu\text{Jy} < \text{S}_{3.6\mu\text{m}}$

Figure 3.17: Percentage of galaxies as a function of position in ELAIS-N1 with $\chi^2 \leq 5$. Panels a) to d) show the effect for galaxies in progressively brighter flux bins. The blue boundary is an artifact of the plotting.

field, satisfying the χ^2 constraint. We attribute this to fainter galaxies typically having smaller signal-to-noise ratios and therefore a smaller χ^2.

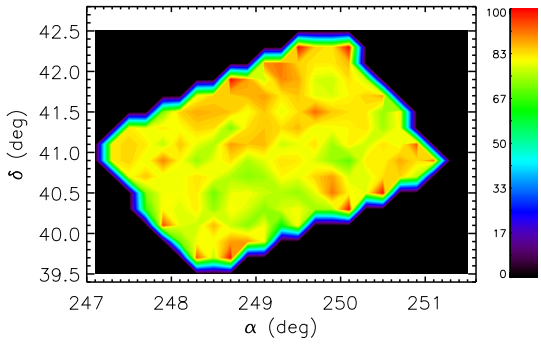
Therefore $\chi^2 \leq 5$ and $n_{band} \geq 5$ may not necessarily be the optimal values to select good photometric redshifts and the optimal values may vary with luminosity and redshift. Figure 3.19 shows how $S_{3.6\mu m}$ and r vary in the stellar mass-redshift plane. These insights indicate a more careful treatment to select "good" photometric redshifts is required.

The clustering analyses outlined in Chapter 5 binned the galaxies into stellar mass-redshift and SFR-redshift *cells*. To ensure accurate photometric redshifts were used in each cell we sought to minimise the fraction of photometric redshift outliers, beyond that achieved by the r-band magnitude cut. We defined an outlier as $dz/(1 + z_{sp}) > 0.1$ at $z = 0$ and > 0.06 at $z = 1$ which corresponded to approximately half a redshift bin. Across

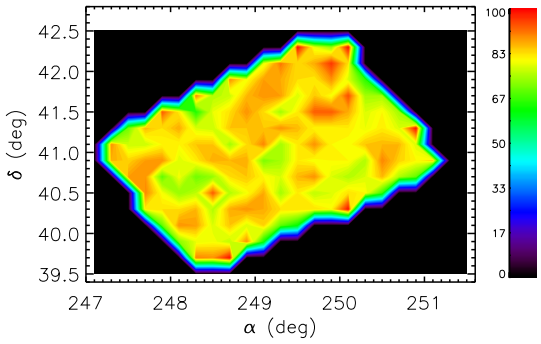


(a) $10\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 20\mu\text{Jy}$

(b) $20\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 40\mu\text{Jy}$

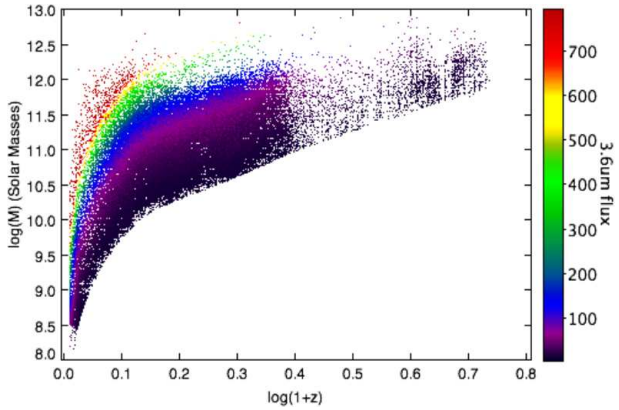


(c) $40\mu\text{Jy} < \text{S}_{3.6\mu\text{m}} < 80\mu\text{Jy}$

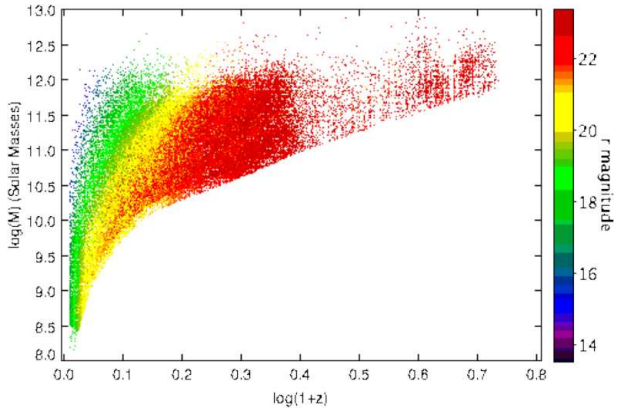


(d) $80\mu\text{Jy} < \text{S}_{3.6\mu\text{m}}$

Figure 3.18: Percentage of galaxies as a function of position in ELAIS-N2 with $\chi^2 \leq 5$. Panels a) to d) show the effect for galaxies in progressively brighter flux bins. The blue boundary is an artifact of the plotting.



(a) $S_{3.6\mu m}$



(b) r

Figure 3.19: Distribution of sources in the stellar mass - redshift plane colour coded by a) $S_{3.6\mu m}$ and b) apparent r-band magnitude.

all fields 5976 sources had spectroscopic redshifts. We searched the NASA/IPAC Extragalactic Database (NED) for sources coincident with the SWIRE fields with spectroscopic redshifts. These sources were cross matched with the SWIRE photometric catalogues using a $1''$ radius. An additional 1293 spectroscopic redshifts were added to the photo-z catalogues.

We investigated how the fraction of outliers in each cell varied with different combinations of n_{band} and χ^2 cuts. We defined cuts to minimise the fraction of outliers. The cuts were based primarily on n_{band} as this translates more clearly to the data as opposed to the χ^2 which has a complex dependancy on the data. We determined the combination of cuts for each cell manually taking into account the number of galaxies rejected and the improvement in the fraction of outliers. For the stellar mass-redshift analysis, above $z = 1$ (cells 16-20) too few galaxies have spectroscopic redshifts to reliably determine the best cuts, therefore $n_{band} \geq 5$ and $\chi^2 \leq 50$ were used. For the SFR-z analysis all cells contained too few galaxies with spectroscopic counterparts, therefore $n_{band} \geq 5$ and $\chi^2 \leq 50$ were also used. See Table 3.6 for the cuts used for each $(M_\star, z)$ cell and the fraction of outliers remaining after these cuts had been applied. How the n_{band} and χ^2 cuts were incorporated into the clustering analyses is outlined in Section 5.2.4.

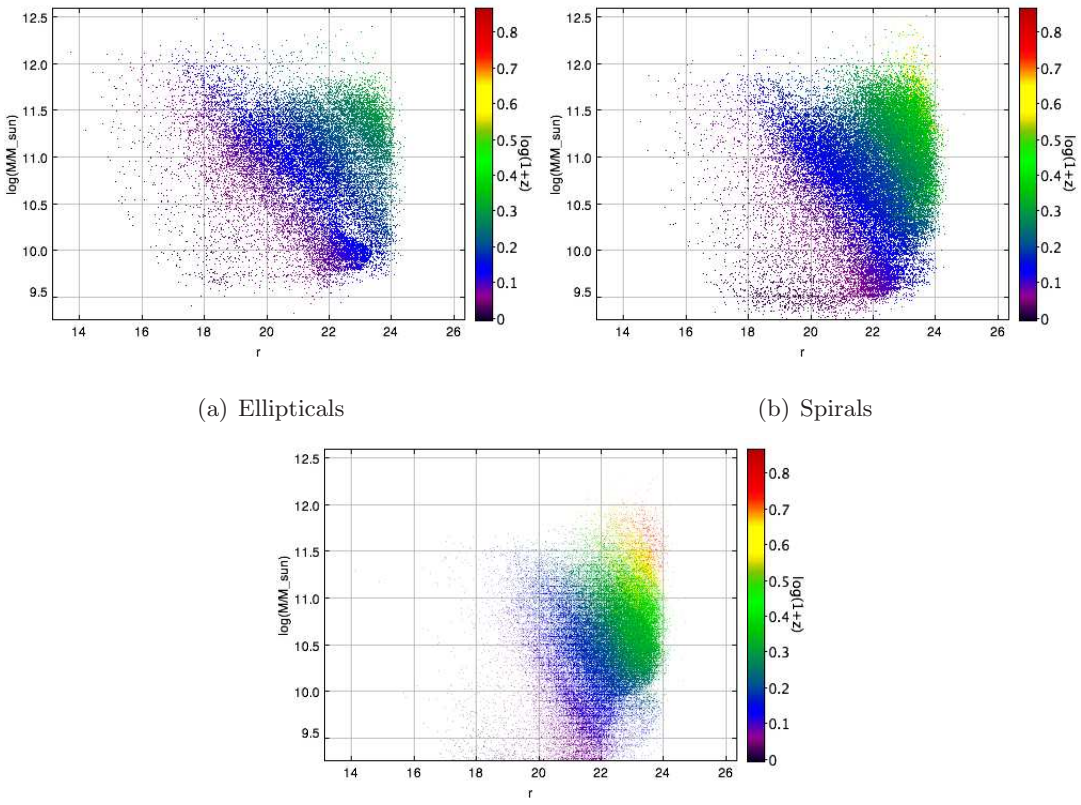
3.8 Stellar Mass and SFR Limits

The r, $S_{3.6\mu\mathrm{m}}$ and $S_{24\mu\mathrm{m}}$ cuts defined in Section 3.7, when applied to the data, produce a homogeneous data set containing good photometric redshifts. However the different templates have different colours and hence will be affected by the magnitude and flux cuts in different ways. Therefore it is important to determine if these cuts preferentially remove more of one type of galaxy than another. Figures 3.20 a) to c) show the stellar mass plotted against r for elliptical, spiral and starburst templates, with the data points colour coded by redshift. These plots indicate that the r cut will produce a stellar mass incompleteness at different stellar masses and redshifts for the different templates. Without determining exactly how the r, $S_{3.6\mu\mathrm{m}}$ and $S_{24\mu\mathrm{m}}$ cuts impose stellar mass and SFR limits on the different types of galaxies it would be difficult to determine whether a variation in clustering strength with stellar mass or SFR or redshift is a real physical trend or due to a variation in the mix of different types of galaxies imposed by these cuts.

Therefore we calculated stellar mass limits (i.e. the lowest mass galaxy which could

Table 3.6: The n_{band} and χ^2 cuts for each $(M_\star, z)$. The fraction of photometric redshifts outliers is also given. The high redshift cells have too few galaxies with spectroscopic redshifts. Therefore the fraction of outliers cannot be determined with any confidence.

Cell	n_{band}	χ^2	Outlier (%)	z	$\log(M_\star/\mathrm{M}_\odot)$
0	6	10	7	0.22	10.25
1	5	10	7	0.24	10.73
2	5	10	8	0.24	11.21
3	6	10	5	0.42	10.38
4	6	10	10	0.41	10.76
5	6	10	10	0.41	11.20
6	5	5	9	0.41	11.70
7	6	10	10	0.58	10.75
8	5	10	11	0.58	11.20
9	5	5	13	0.58	11.60
10	5	5	14	0.80	10.85
11	6	98	15	0.83	11.23
12	6	10	10	0.84	11.61
13	6	5	15	0.95	10.93
14	6	5	18	0.98	11.26
15	5	5	17	1.00	11.63
16	5	50	-	1.14	10.95
17	5	50	-	1.18	11.32
18	5	50	-	1.19	11.66
19	5	50	-	1.37	11.31
20	5	50	-	1.40	11.65



(a) Ellipticals

(b) Spirals

(c) Starburst

Figure 3.20: The stellar mass plotted against r for a) elliptical templates (E), b) spiral templates (Sab-Scd) and c) starburst templates (Sdm-sb). Points are colour coded by redshift.

be observed) and SFR limits imposed by the $S_{3.6\mu\mathrm{m}}$, $S_{24\mu\mathrm{m}}$ and r-band limits for E, Sab, Sbc, Scd, Sdm and sb template galaxies as follows. The mass-to-light ratio used in Rowan-Robinson et al.(2008) is a simple one that varies with type and redshift,

$$\frac{(M_*/M_\odot)}{(\nu L_\nu(3.6)/L_\odot)(t)} = \frac{50}{(a + 1.17(t/t_0)^{-0.6})} \tag{3.20}$$

where t_0 is the present epoch and a = 0.15, 0.08, 0.61, 0.26, 1.44, 0.70 for SED types E, Sab, Sbc, Scd, Sdm and sb, respectively. Therefore, $M_\star \propto r$ and similarly $M_\star \propto S_{3.6\mu m}$ so the stellar mass limits imposed by the $S_{3.6\mu m}$ and r band cuts are given by,

$$\frac{M_{\star,lim}}{M_\star} = \frac{S_{3.6\mu m,lim}}{S_{3.6\mu m}} \tag{3.21}$$

$$\log(M_{\star,lim}) = \log\left(\frac{S_{3.6\mu m,lim}}{S_{3.6\mu m}}\right) + \log(M_\star) \tag{3.22}$$

and similarly for r,

$$\frac{M_{\star,lim}}{M_\star} = \frac{S_{r,lim}}{S_r} \tag{3.23}$$

$$-2.5\log(M_{\star,lim}) = -2.5\log(S_{r,lim}) + 2.5\log(S_r) - 2.5\log(M_\star) \tag{3.24}$$

$$\log(M_{\star,lim}) = -0.4(r_{lim} - r) + \log(M_\star) \tag{3.25}$$

which also applies for u,g,i,z optical bands. We defined the magnitude limits for the u,g,i,z bands to be the turnover in the number counts, see Figures 3.21.

The SFR was estimated by Rowan-Robinson et al.(2008) using the conversion from 60μm luminosity of Rowan-Robinson et al.(1997), Rowan-Robinson(2001),

$$SFR(\mathrm{M_\odot yr^{-1}}) = 2.2\epsilon^{-1}10^{-10}\frac{\mathrm{L_{60}}}{\mathrm{L_\odot}} \tag{3.26}$$

where ϵ is the fraction of UV light absorbed by dust and taken to be 2/3 and the bolometric corrections at $60\mu m$ needed to convert L_{ir} to L_{60}, are 3.48, 1.67 and 1.43 for cirrus, M82 and A220 templates, respectively. Therefore, $SFR \propto r$ and similarly $SFR \propto S_{24\mu m}$, so we calculated the SFR limits imposed by the $S_{24\mu m}$ and r band cuts in the same way as the stellar mass limits were calculated giving,

$$\log(SFR_{lim}) = \log\left(\frac{S_{24\mu m,lim}}{S_{24\mu m}}\right) + \log(SFR) \tag{3.27}$$

and

$$\log(SFR_{lim}) = -0.4(r_{lim} - r) + \log(SFR) \tag{3.28}$$

Figure 3.22 shows the stellar mass limits imposed due to the $S_{3.6\mu\mathrm{m}}$ and r-band cuts. However we should remember that the r band limit is not a sharp limit as it depends on

Table 3.7: The magnitude and flux cuts used for each waveband to calculate stellar mass and SFR limits.

u	g	r	i	z	3.6μm μJy	4.5μm μJy	24μm μJy
22.7	23.8	23.2	22.4	21.2	10	15	400

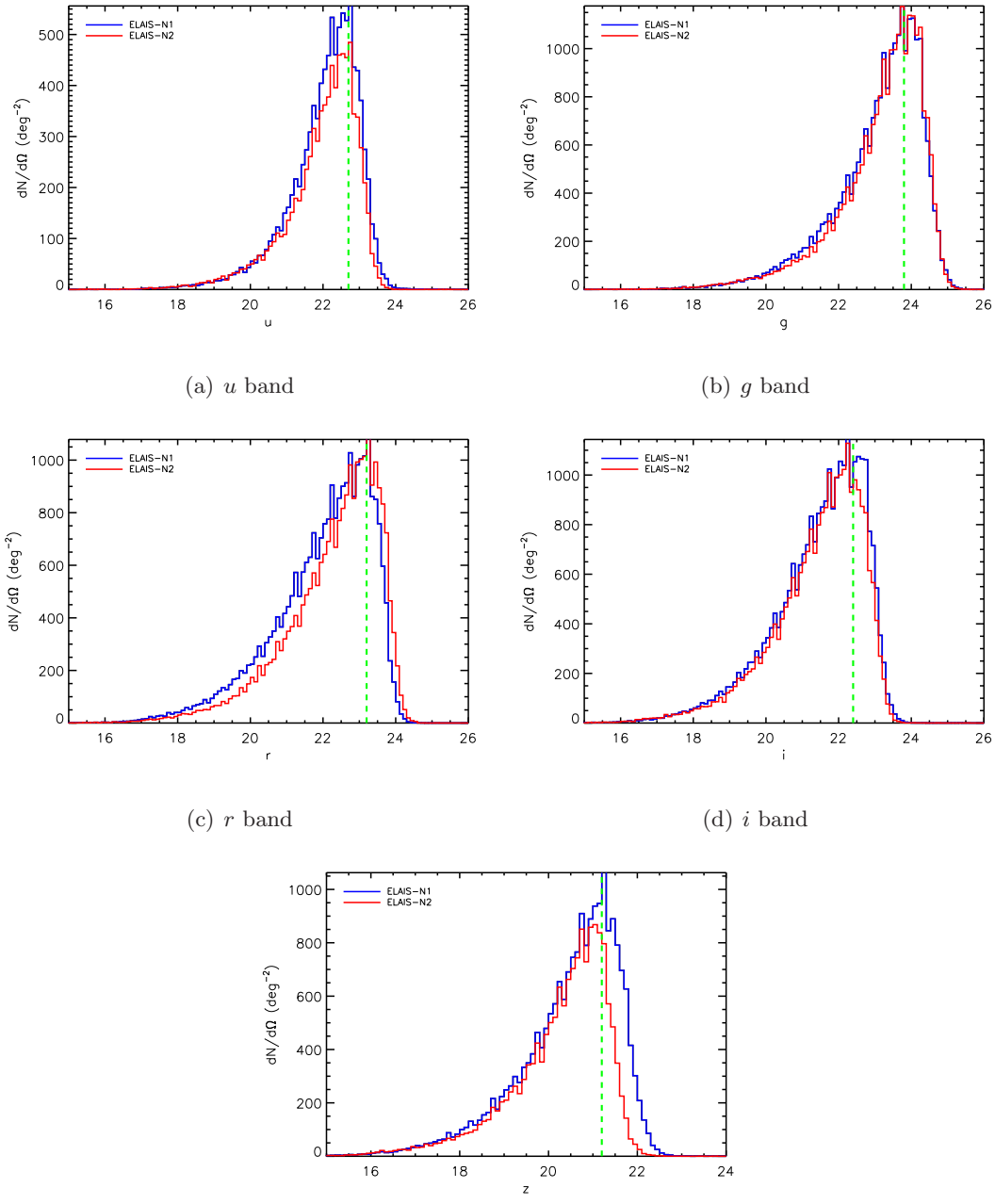
both $M_\star$ and SFR. The $S_{3.6\mu\mathrm{m}}$ limit did not bias against any type of galaxy as expected as the galaxies are selected on stellar mass. However, the r-band limit was strongly biased against early-type galaxies at $z > 0.5$ as can be seen from the spread of the stellar mass limits, producing what we term a stellar mass incompleteness. It also introduced a stellar mass incompleteness for spiral galaxies. These galaxies are massive enough to enter the mass selected samples, but too faint to fulfill the r_{cut} limit. Alternatively their M/L ratios are large and they tend to have redder colours.

The clustering analysis (Chapter 5) separately measured the clustering of elliptical (E), spiral (Sab-Scd) and starburst (Sdm-sb) galaxies. Therefore any change in the relative mix of spectral types due to the cuts was minimal. Any high redshift stellar mass-redshift cells which where strongly affected by the biasing or suffered from a stellar mass incompleteness due to r_{cut} were removed from the analysis. Figure 3.23 shows the SFR limits imposed by the $S_{24\mu\mathrm{m}}$ and r-band limits. The SFR limit imposed by the $S_{24\mu\mathrm{m}}$ dominates at all redshifts and was not biased against certain types of galaxies.

The stellar mass and SFR limits are not just specific to SWIRE. Different photometric limits in different surveys will simply act to shift the stellar mass or SFR limit curves up or down. It is interesting to note the r-band needs to be much deeper to obtain a truly homogeneous data set out to high redshift.

3.9 Summary

In this chapter a brief introduction to IR astronomy was given. The SWIRE survey was introduced along with the SWIRE photometric redshift catalogue (Rowan-Robinson et al.(2008)) which also utilizes the INTWFS survey. SWIRE completeness curves were shown and explained and INTWFS completeness curves were calculated. The selection effects in the photometric redshift catalogue resulting from the selection effects of the SWIRE and INTWFS surveys have been thoroughly investigated. For the stellar mass-



(a) u band

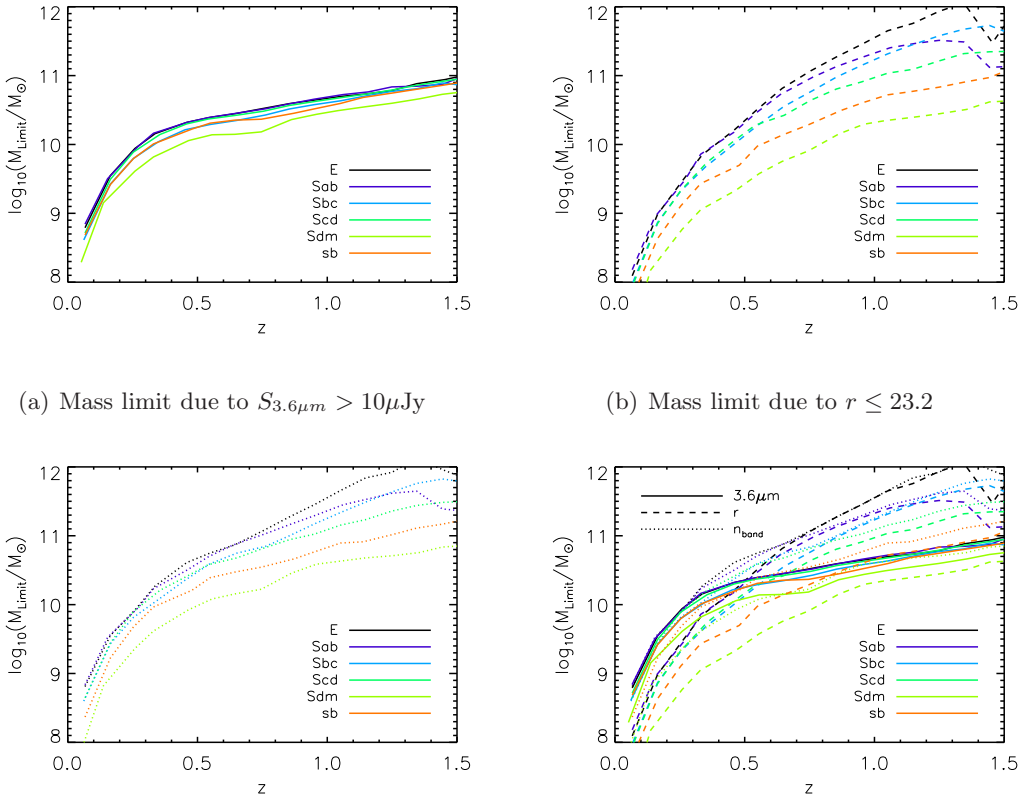
(b) g band

(c) r band

(d) i band

(e) z band

Figure 3.21: Number counts in u,g,r,i,z bands for ELAIS-N1 (blue) and ELAIS-N2 (red). The green dashed line indicates the point at which the number counts turn over which we defined to be our magnitude limit for calculating the stellar mass and SFR limits. In panel c), the r-band, the green dashed line shows r_{cut} determined in Section 3.7.3.



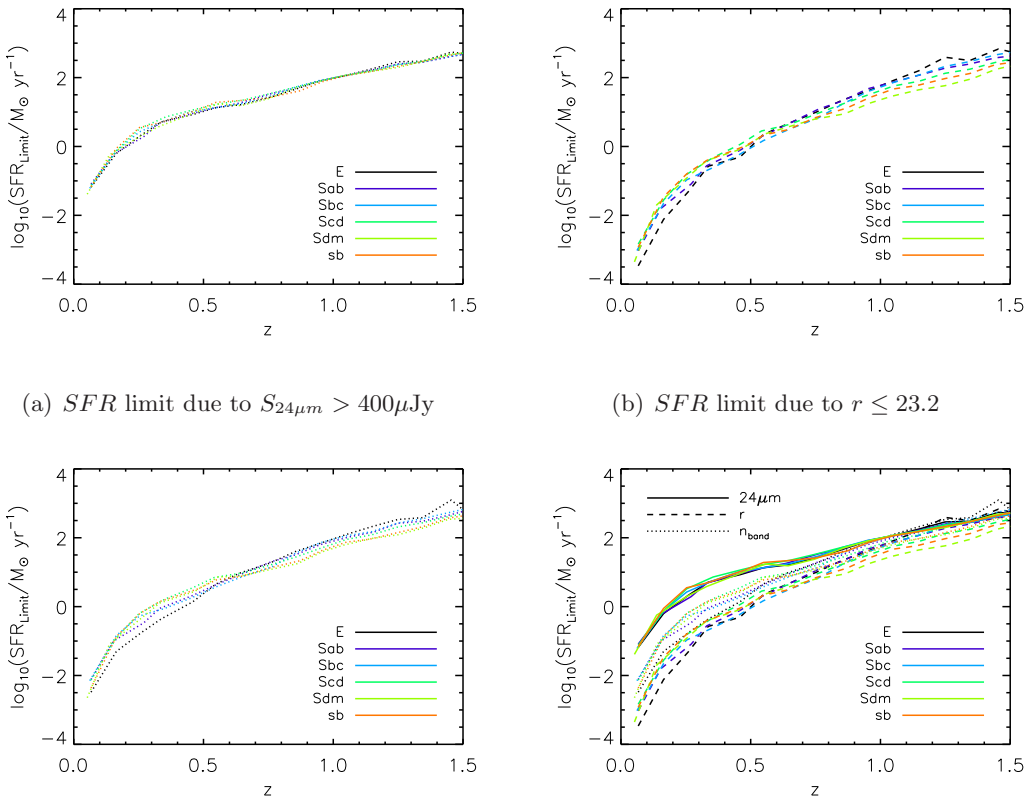
(a) Mass limit due to $S_{3.6\mu m} > 10\mu$Jy

(b) Mass limit due to $r \leq 23.2$

(c) Mass limit due to $n_{band} \geq 5$

(d) All three limits

Figure 3.22: The stellar mass limits resulting from the 3.6μm flux cut of 10μJy (solid lines), $r \leq 23.2$ (dashed lines) and $n_{band} \geq 5$ (dotted lines). Galaxy types E, Sab, Sbc, Scd, Sdm and sb are shown, the colours for which are shown in the legend.



(a) SFR limit due to $S_{24\mu m} > 400\mu$Jy

(b) SFR limit due to $r \leq 23.2$

(c) SFR limit due to $n_{band} \geq 5$

(d) All three limits

Figure 3.23: The SFR limits resulting from the 24μm flux cut of 400μJy (solid lines), $r \leq 23.2$ (dashed lines) and $n_{band} \geq 5$ (dotted lines). Galaxy types E, Sab, Sbc, Scd, Sdm and sb are shown, the colours for which are shown in the legend.

redshift (or $SFR-z$) analysis 3.6μm flux (24μm), r-band magnitude, 3.6μm IRAC (24μm MIPS) coverage, n_{band} and χ^2 cuts have been defined. Stellar mass and SFR incompleteness resulting from the flux and magnitude cuts was investigated. Binary masks were made to apply the coverage cuts and remove foreground stars from the fields. The masks also contain an optical component to remove regions of the field which have a low optical completeness. The cuts and the masks were made to provide a reliable homogeneous data set which still retains a sufficiently large number of galaxies which have good photometric redshifts. How they are incorporated into the clustering analyses is outlined in Chapter 5.

Chapter 4

Clustering in the Millennium Simulation

4.1 Introduction

In this chapter there is a brief introduction to semi-analytic models and the Millennium Simulation. The spatial correlation functions of dark matter halo catalogues from the Millennium Simulation are measured for a range of masses and redshifts. Halo merger trees are used to identify progenitors of redshift zero haloes at earlier times. The clustering of such haloes is measured at select redshifts in order to map the evolution of the clustering of dark matter haloes. The semi-analytic model of De Lucia & Blaizot(2007) is used to measure the clustering of synthetic galaxies selected on stellar mass and redshift.

4.2 Cosmological Simulations

Modern day astronomy can be sub-divided in many different ways but arguably the most important distinction is between observational astronomy and theoretical modeling using computer simulations. Observations help constrain models and computer simulations provide predictions for and comparisons to observations. Together they complement each other to build up a picture of how the Universe works.

Over the past 50 years the processing power of computers has continually increased following Moore's law - computing power doubles every 18 months. Cosmological simulations are limited in spatial extent and mass resolution by the processing power available.

The more powerful a computer is, the more simulation particles can be used to trace the mass in the Universe and/or smaller mass resolution can be achieved. There are many different types of simulations which utilize todays powerful computers but this chapter will focus on semi-analytic models.

4.3 Semi-Analytical models

The ideal cosmological simulation would include dark and baryonic matter and be large enough to encompass all the different facets of large scale structure such as filaments, voids, clusters and superclusters as well as resolve small-scale processes associated with galaxy formation and evolution such as star formation, supernovae, AGN feedback, stellar winds etc. The processing power to run such a simulation in a realistic timescale does not currently exist, neither does the detailed understanding of the small-scale processes.

Semi-analytical modeling provides a way to develop models which are spatially large and which allow galaxy evolution to be monitored at siginificantly reduced computational cost. The idea was originally coined by White & Rees(1978) whereby the dark matter structure of the Universe is modeled using a Press-Schechter (Press & Schechter 1974), Monte-Carlo or N-body approach. The processes associated with galaxy formation and evolution are then added in as parameterized equations. This gives a simplified treatment of processes such as gas cooling and star formation but still retains a physical basis.

The first proper semi-analytical model was made by White & Frenk(1991). An important development whereby the formation and evolution of galaxies were simulated in evolving dark matter haloes was made by Kauffmann et al.(1993) and Cole et al.(1994). These advancements paved the way for the cosmological semi-analytic models that are familiar today such as the Millennium Simulation, Springel et al.(2006). See Baugh(2006) for an excellent review on semi-analytic models.

4.4 The Millennium Simulation

The Millenium Simulation is based on a large N-body simulation. It is comprised of 10^{10} dark matter particles which are evolved from $z = 127$ to the present day in a cubic region of space 500 h^{-1}Mpc on a side. It has a mass resolution of 8.6×10^8 $h^{-1}\mathrm{M}_\odot$. *Merger trees* are generated to determine the merger history of each *root* halo. Working backwards in time the merger tree is built up step-by-step by identifying the haloes at the previous

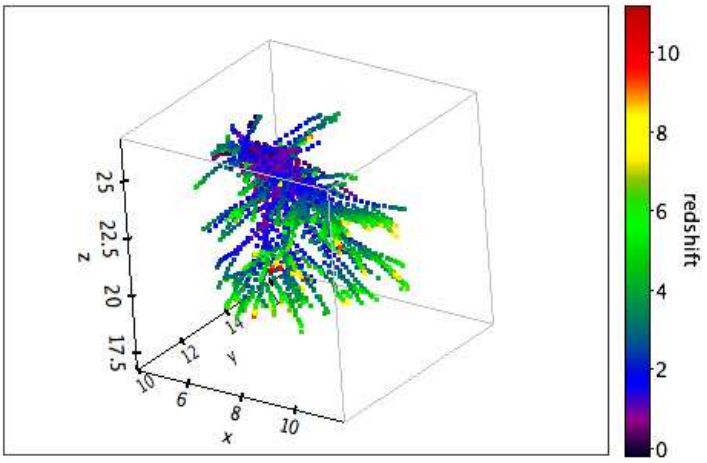


Figure 4.1: The x, y and z cartesian coordinates in Mpc/h of the constituent haloes of the merger tree of a redshift zero halo. The haloes are colour coded by redshift indicated by the colour bar. Over time the haloes have come together under the force of gravity until they are tightly clustered at $z = 0$.

redshifts from which the present halo formed. Figures 4.1 and 4.2 show an example of a merger tree. Semi-analytic models are implemented into the simulation to follow gas, star and supermassive blackhole processes within the merger history trees of the dark matter haloes and their associated sub-haloes. Galaxy and dark matter halo catalogues are output at 63 distinct redshifts, termed SNAPNUMS, between $0 < z < 20$. See Springel et al.(2006) for a full description of the simulation. Figure 4.3 shows a region of the redshift zero dark matter catalogue. The large scale structure of the Universe is clearly visible indicating the Millennium Simulation is an excellent tool for investigating large scale clustering.

4.4.1 The Clustering of Dark Matter Haloes

The dark matter halo catalogues from the Millennium Simulation were used to measure the large scale inter-halo clustering of dark matter haloes over a range of redshifts. The catalogues are publicly available and the database can be accessed via an SQL based web interface [1]. Two analyses were performed:

- *Fixed mass clustering analysis:* The clustering of haloes of masses $\log(M_{halo}/\mathrm{M}_{\odot})$ =10.5, 11.0, 11.5, 12.0, 12.5, 13.0 and 13.5 were measured in comoving coordinates at select

[1] http://www.g-vo.org/Millennium

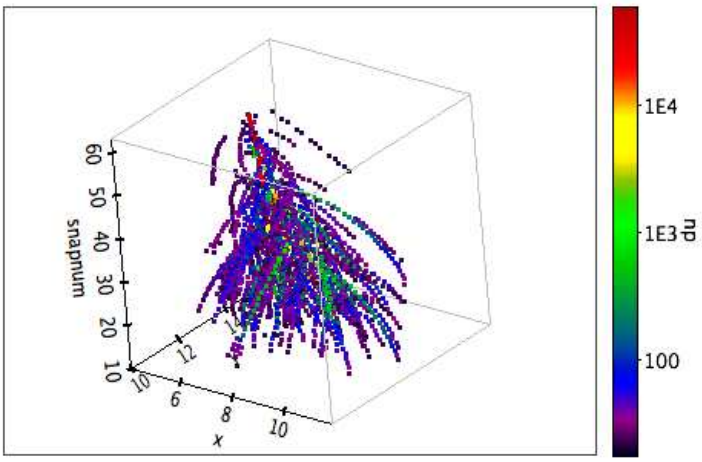


Figure 4.2: The x and y positions in Mpc/h as a function of SNAPNUM (redshift) of the constituent haloes of the merger tree of a redshift zero halo. The haloes are colour coded by the number of simulation particles comprising each halo, indicated by the colour bar. Over time the haloes have come together under the force of gravity to produce more massive haloes as they have merged.

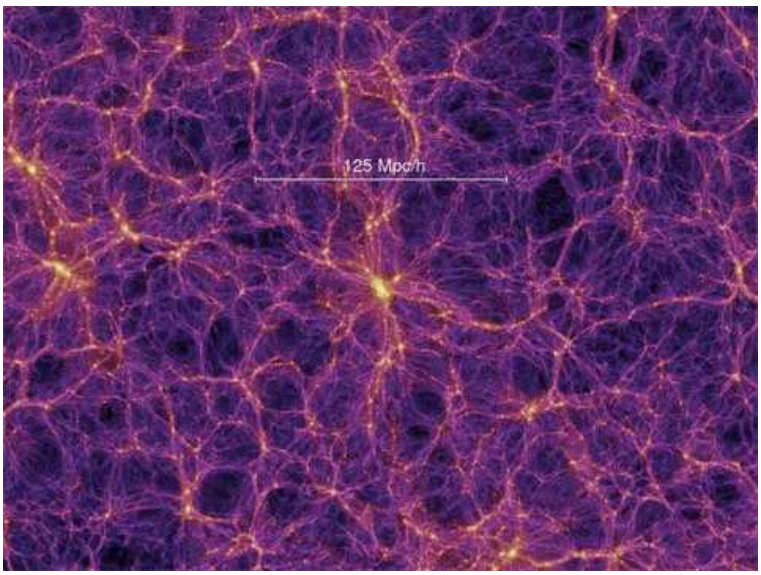


Figure 4.3: The dark matter distribution as seen in the Millennium Simulation, Springel et. al 2005. The large scale structure of the Universe is clearly visible.

redshifts over $0 < z < 1.5$, i.e. we selected haloes of the same mass at each redshift to see how the clustering strength of such objects varied with redshift.

- *Evolution of clustering analysis:* Haloes of masses $\log(M_{halo}/\mathrm{M}_{\odot})$ =11.0, 11.5, 12.0, 12.5, 13.0, 13.5 and 14.0 were selected at $z = 0$ and the progenitors of these haloes were also selected at certain redshifts out to $z = 1.5$. The clustering of the $z = 0$ haloes of each mass was measured as was the clustering of the progenitors in each mass range at higher redshifts, i.e. the same haloes were selected at each redshift enabling the evolution of the clustering of the dark matter haloes to be measured.

The catalogues for each analysis were obtained as follows.

4.4.2 Fixed Mass Catalogues

We obtained catalogues containing information about the spatial positioning of dark matter haloes from the Millennium Simulation database to investigate how dark matter haloes of the same mass cluster at different times. Here we use M_{hz} to denote the halo mass at redshift z. The spatial positioning of galaxies in the simulation cube are given by 3 cartesian coordinates, x, y, z. We obtained haloes of mass $\log(M_{hz}/\mathrm{M}_{\odot})$=10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5 at z =0, 0.12, 0.32, 0.51, 0.69, 0.91 and 1.5 as follows. We determined the masses of haloes in the database by counting how many simulation particles they were comprised of, knowing each particle has a mass of $8.6 \times 10^8\ h^{-1}\mathrm{M}_{\odot}$. We selected haloes of each mass at each redshift based on the number of particles. The SNAPNUMS and corresponding redshifts at which the catalogues were obtained are shown in Table 4.1. We calculated the spatial correlation function for each mass range at each redshift as outlined in Section 4.4.4

4.4.3 Evolution Catalogues

For the evolution analysis we used M_{h0} to denote the halo mass at $z = 0$. At redshift zero we selected haloes of mass $\log(M_{h0}/\mathrm{M}_{\odot})$=11.0, 11.5, 12.0, 12.5, 13.0, 13.5 and 14.0 in the same manner as described in the fixed mass analysis. We traced the progenitors of these haloes, at earlier redshifts, through the simulation to $z = 1.6$. We did this using the merger tree for each halo. We only required haloes located on the main branch of the merger tree as our analysis was only concerned with measuring the large scale inter-halo clustering. In the database every halo in the same merger tree has the same LASTPROGENITORID which

Table 4.1: SNAPNUMS and the corresponding redshifts at which dark matter halo catalogues were obtained from the Millennium Simulation.

Snapnum	Redshift
63	0.00
58	0.12
52	0.32
48	0.51
45	0.69
52	0.91
36	1.50

is an ID number assigned to the halo on the main branch of the merger tree at the earliest SNAPNUM at which the tree began to form. The HALOID assigned to the halo on the main branch of the merger tree decreases by one at every subsequent SNAPNUM. Therefore we identified the halo on the main branch at each SNAPNUM using the following condition, HALOID+SNAPNUM=LASTPROGENITORID, which is a constant for the halo on the main branch at every SNAPNUM. We extracted the halo on the main branch at redshifts z=0.0, 0.12, 0.32, 0.62, 1.08 and 1.63 corresponding to SNAPNUM=63, 58, 52, 46, 40 and 35. We were selecting the same haloes at each redshift, the masses of which were smaller at higher redshifts. The spatial correlation function was then calculated for each halo catalogue at each redshift as outlined in Section 4.4.4

4.4.4 Measuring the Clustering of Dark Matter Haloes

We used the same method to measure the clustering of the haloes for each analysis. The positions of the haloes within the cube were specified in x,y,z cartesian coordinates in units of h^{-1}Mpc. A catalogue of 500,000 random sources with a Poisson distribution was created for each halo catalogue which matched the spatial extent of the simulation cube. We split the halo and random catalogues into 8 equal volumes to determine errors on the correlation function using a jackknife resampling technique. Each octant was removed in turn and the number of data-data (DD), data-random (DR) and random-random (RR) pairs separated by a distance r were counted for the remaining subsamples using the NPT (Gray et al.(2001)) pair counting code.

We calculated the spatial correlation function (Equation 2.5) for halo catalogue i for jackknife sample j, at large scales of $4 - 40\ h^{-1}$Mpc, using the estimator given by Landy & Szalay(1993),

$$\xi(r)_{ij} = \frac{DD_{ij} - 2DR_{ij} + RR_{ij}}{RR_{ij}} \tag{4.1}$$

For catalogue i the N measures of $\xi(r)$ were averaged together using,

$$\xi_i(r) = \sum_{j=0}^{j=N} \xi_{ij}(r)/N \tag{4.2}$$

The error on $\xi_i(r)$ was calculated using,

$$\sigma_{\xi_i(r)} = \sqrt{\frac{N-1}{N} \frac{\sum_j^N (\xi_{ij}(r) - \xi_i(r))^2}{N}} \tag{4.3}$$

A power law of the form of Equation 2.5 was fitted to the spatial correlation function where γ was kept fixed at 1.8. The amplitude of the power law gave a measure of the comoving spatial clustering strength r_0.

4.4.5 The Clustering of Haloes of Fixed Mass

The spatial correlation functions were reliably measured for all masses and redshifts probed. There was no variation in γ over the spatial scales probed. Figure 4.4 shows the spatial correlation functions for the fixed mass analysis.

Figure 4.5 shows the comoving clustering strength r_0 for each halo mass as a function of redshift. Haloes with $\log(M_{hz}/\mathrm{M}_\odot) \leq 12.0$ showed a small decrease in clustering strength with redshift, an effect which was more pronounced for the lowest mass haloes. Conversely haloes with $\log(M_{hz}/\mathrm{M}_\odot) \geq 12.5$ showed an increase in clustering strength at high redshift. A variation which was more pronounced for the highest mass haloes.

These trends can be explained as the result of rarer haloes clustering more strongly because they tend to be found closer together (e.g. Kaiser(1984), Efstathiou et al.(1988)). Massive haloes trace some of the highest peaks in the density field. At high redshift these peaks are very rare whereas they are more frequent at low redshift as more time has passed allowing many of such peaks to form through hierarchical mergers. For example, there were 3331 $\log(M_{hz}/\mathrm{M}_\odot) = 13.5$ haloes at $z = 0$ and 890 at $z = 1.5$. The same reasoning applies to the low mass haloes, however they were rarer at $z = 0$ than $z = 1.5$ and the relative difference was less extreme. For example, there were 54,589 $\log(M_{hz}/\mathrm{M}_\odot) = 11.0$ haloes at $z = 0$ and 59,738 at $z = 1.5$.

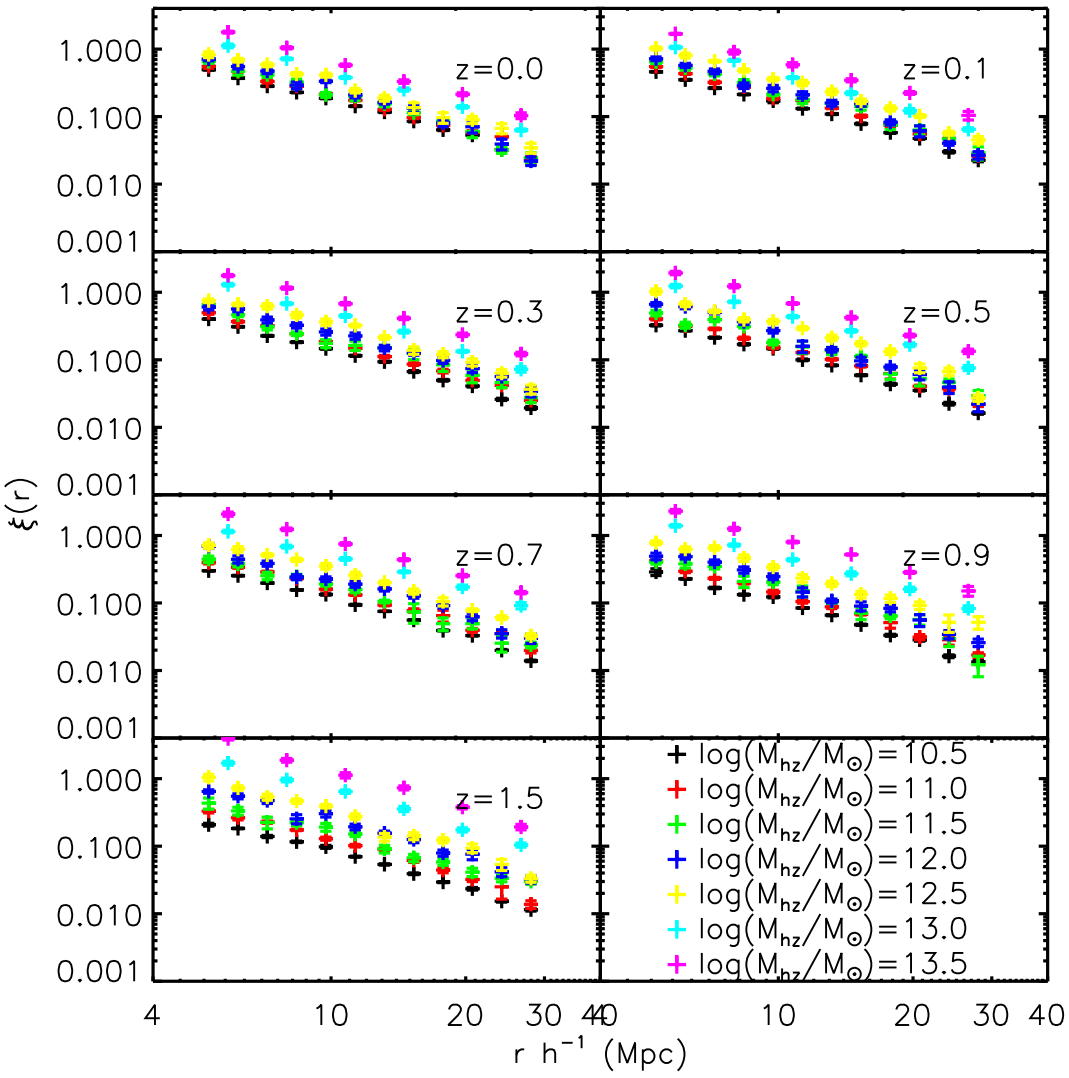


Figure 4.4: The spatial correlation function, $\xi(r)$, for dark matter haloes of mass $\log(M_{hz}/\mathrm{M}_{\odot})$=10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5 at z =0.1, 0.3, 0.5, 0.7, 0.9 and 1.5. Each panel shows the correlation functions for every mass range at a different redshift, indicated in the panel. Power law fits are omitted for clarity.

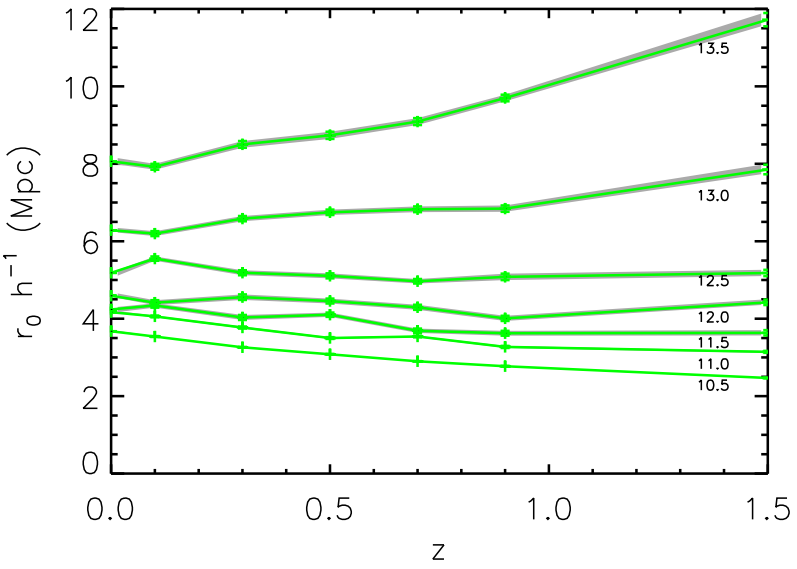


Figure 4.5: The comoving clustering strength r_0 for haloes of mass $\log(M_{hz}/\mathrm{M}_\odot)$=10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5 over $0 < z < 1.5$. The grey regions show the jackknife errors on the clustering strength.

The rate of change of clustering strength with halo mass (at a given redshift),

$$\left(\frac{dr_0}{d\log(M_{hz})}\right)_z (M_{hz}) \tag{4.4}$$

by looking at Figure 4.5, was larger for more massive haloes ($12.0 < \log(M_{hz}/\mathrm{M}_\odot) < 14.0$) and smaller for less massive haloes ($10.5 < \log(M_{hz}/\mathrm{M}_\odot) \leq 12.0$). This is because haloes are biased tracers of the dark matter distribution. Massive haloes are more biased than less massive haloes and hence they cluster more strongly, an effect which increases with increasing mass.

We should also consider that recent numerical simulations have shown how halo clustering depends on properties such as formation redshift, halo concentration, sub-halo composition and halo spin (Gao et al.(2005), Wechsler et al.(2006), Harker et al.(2006), Bett et al.(2007), Wetzel et al.(2007), Jing et al.(2007), Espino-Briones et al.(2007), Zhu et al.(2006), Croton et al.(2007)). Our analysis probed the large scale inter-halo clustering, hence it was not sensitive to the sub-halo structure. The dependency on formation redshift is also known as *assembly bias* (Croton et al.(2007)) where the clustering of haloes is shown to be dependent on their assembly history. Gao et al.(2005), Zhu et al.(2006),

Wechsler et al.(2006) and Croton et al.(2007) showed that low mass haloes that assembled at high redshift were more clustered than those of the same mass that assembled more recently. Conversely they found high mass haloes at high redshift to be less clustered than those that formed more recently.

Bett et al.(2007) studied the clustering dependence on halo spin in the Millennium Simulation and found haloes with higher spin were more strongly clustered speculating it might be because haloes in more clustered environments are more likely to experience stronger tidal forces, leading to more coherent rotation. This effect was only observed for high mass ($M_{halo} \gtrsim 10^{12} \mathrm{M}_{\odot}$) haloes at $z = 0$ and was not investigated at higher redshifts. Halo clustering dependence on concentration was investigated by Wetzel et al.(2007), Angulo et al.(2008) and Wechsler et al.(2006). Low mass haloes with high concentration were found to be more clustered than low mass halos of low concentration, an effect which was stronger with decreasing halo mass. For massive haloes the trend was reversed and halos of low concentration were more strongly clustered than their high-concentration counterparts.

Our halo catalogues did not contain enough information about the properties of the haloes to conclusively determine how each of the halo properties discussed above affected the clustering strengths. A detailed analysis would be required to investigate the effects of each property individually.

4.4.6 The Evolution of the Clustering of Haloes

The spatial correlation functions were reliably measured for all masses and redshifts probed. There was no variation in γ. Figure 4.6 shows the spatial correlation functions for the $z = 0$ haloes and their progenitors at earlier redshifts, in each mass range.

The clustering strength of the progenitors of the redshift zero haloes decreased with increasing redshift as expected. Figure 4.7 shows the evolution of the comoving clustering strength, r_0, for the range of halo masses. Over time haloes merge under the influence of gravity producing more massive haloes which exert a stronger gravitational force on surrounding haloes eventually making them more clustered. Our results agree with findings from other simulations which looked at the evolution of the clustering of dark matter haloes e.g. Weinberg et al.(2004).

The rate of change of clustering strength with M_{h0} was small in the range $11.0 < \log(M_{h0}/\mathrm{M}_{\odot}) < 12.0$ but much larger in the range $12.0 < \log(M_{h0}/\mathrm{M}_{\odot}) < 13.0$. This was

because the most massive haloes are much rarer and are more biassed tracers of the dark matter distribution, an affect which increases with mass, as was discussed in Section 4.4.5.

4.5 Stellar Mass Clustering of the Semi-Analytic Catalogues

We used the semi-analytic model of De Lucia & Blaizot(2007) in this analysis to measure the clustering of synthetic galaxies selected on stellar mass over a range of redshifts. We measured the spatial correlation function for each stellar mass-redshift synthetic catalogue to determine r_0 in comoving coordinates.

De Lucia & Blaizot(2007) implemented their semi-analytic model into the N-body component of the Millennium Simulation following the technique of Springel et al.(2001). The techniques governing the physical processes of galaxy evolution used by De Lucia & Blaizot(2007) were built upon those developed by earlier works such as Kauffmann et al.(1999), Springel et al.(2001), De Lucia et al.(2004), Springel et al.(2005) and Croton et al.(2006). One of the key ingredients was the inclusion of AGN feedback which suppresses star formation in galaxies with large black holes. They tuned their model to jointly fit the luminosity, colour and morphology distribution of low-redshift galaxies. For full details see De Lucia & Blaizot(2007).

4.5.1 Synthetic Galaxy Catalogues

Catalogues containing the x, y, z cartesian coordinates of the galaxies were obtained for a range of stellar masses and redshifts using the same SQL based web interface as was used to get the dark matter halo catalogues. Each galaxy in the model had a stellar mass estimate in units of solar masses. The stellar masses used were $\log(M_\star/\mathrm{M}_\odot)$=9.3, 9.7, 10.2, 10.8 and 11.2. The redshifts used were z =0.09, 0.32, 0.51, 0.69, 0.91 and 1.28. The two lowest stellar mass bins were only obtained at $z = 0.09$ and the two highest stellar mass bins were only obtained at $z \geq 0.32$, see Table 4.2 for details of each catalogue.

The spatial correlation function was measured for the galaxies in each catalogue over $6 < r < 28\ h^{-1}$Mpc using the same method as that used for the dark matter halo catalogues outlined in Section 4.4.4.

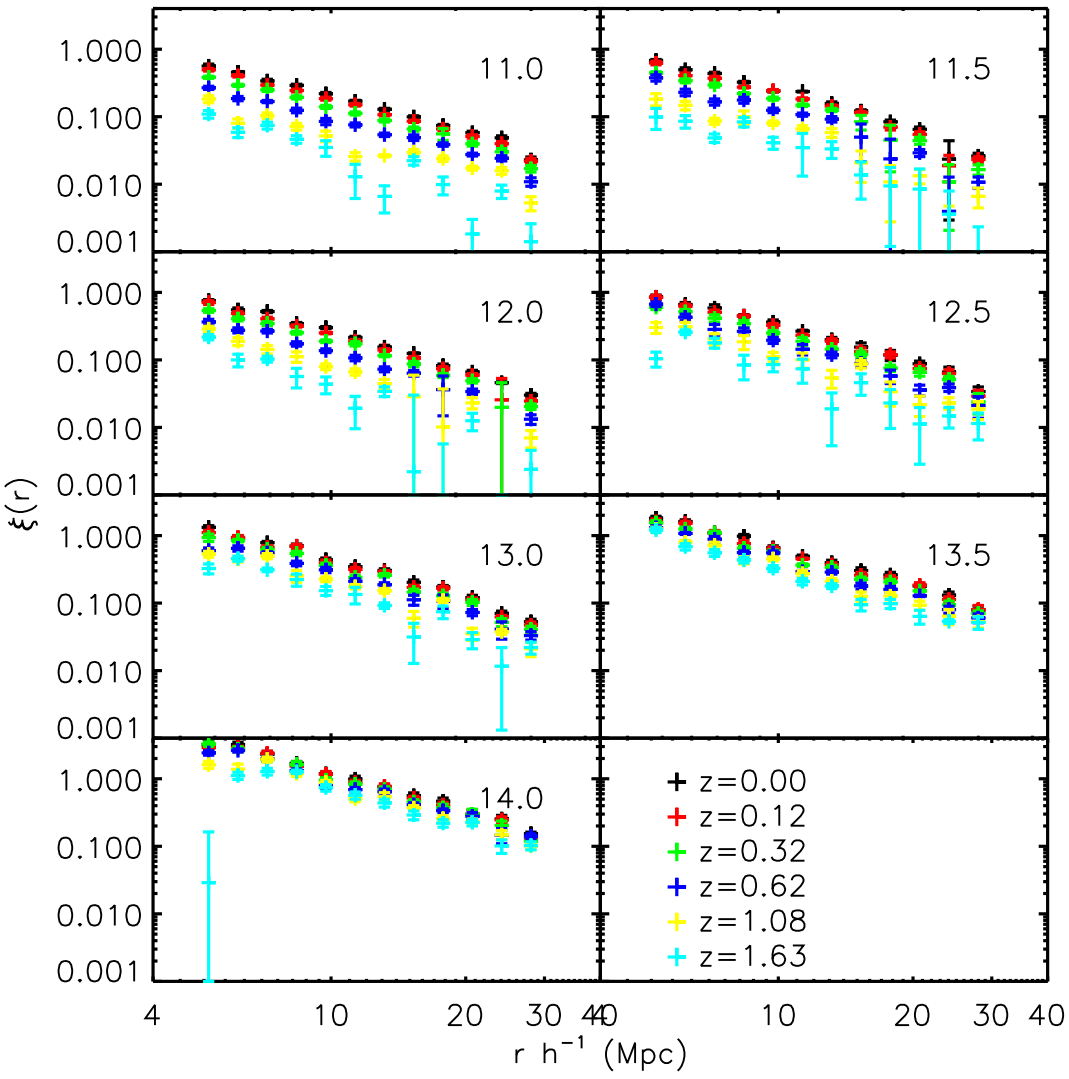


Figure 4.6: The spatial correlation function, $\xi(r)$, for dark matter haloes of mass $\log(M_{h0}/\mathrm{M}_{\odot}) = 11.0$, 11.5, 12.0, 12.5, 13.0, 13.5 and 14.0 and their progenitors at $z =$ 0.12,0.32,0.62,1.08 and 1.63. Each panel shows the correlation functions for one mass range, indicated by the number in the top right hand corner which is the logarithm in solar masses, at every redshift. Power law fits are omitted for clarity.

Table 4.2: The redshift, snapnum, stellar mass and comoving spatial clustering strength for each catalogue obtained from Delucia et. al. 2007 semi-analytic model.

Catalogue	Redshift	SNAPNUM	$\log(M_\star/M_\odot)$	$r_0\ h^{-1}$Mpc
1	0.09	59	9.3	6.16 ± 0.04
2	0.09	59	9.7	6.19 ± 0.04
3	0.09	59	10.2	6.16 ± 0.05
4	0.32	52	10.2	5.83 ± 0.06
5	0.32	52	10.8	6.65 ± 0.06
6	0.32	52	11.2	7.48 ± 0.06
7	0.51	48	10.2	5.76 ± 0.04
8	0.51	48	10.8	6.14 ± 0.06
9	0.51	48	11.2	7.42 ± 0.07
10	0.69	45	10.8	6.30 ± 0.08
11	0.69	45	11.2	8.11 ± 0.05
12	0.91	42	10.8	6.27 ± 0.03
13	0.91	42	11.2	8.58 ± 0.07
14	1.08	40	10.8	6.29 ± 0.03
15	1.08	40	11.2	9.31 ± 0.09

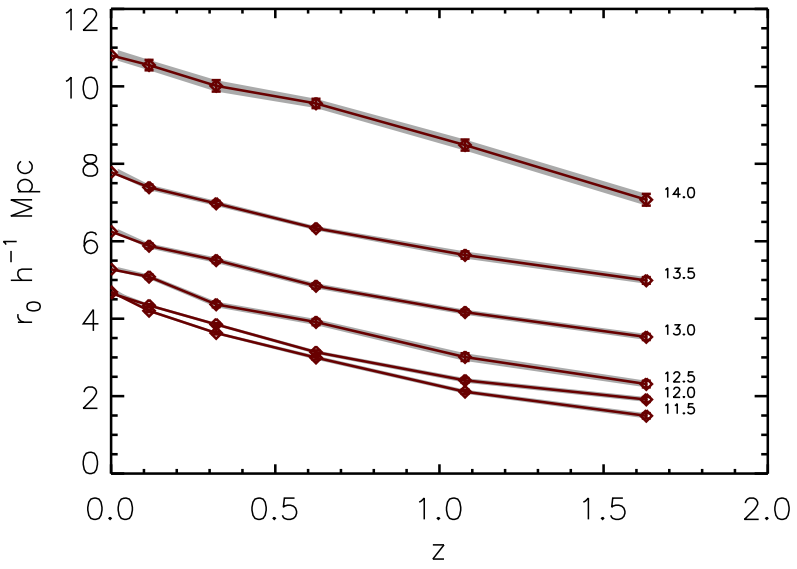


Figure 4.7: The evolution of the comoving clustering strength of the main progenitors of $\log(M_{h0}/\mathrm{M}_{\odot})$=10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5 and 14.0 haloes over $0 < z < 1.5$. The grey regions show the projected jackknife errors on the clustering strength.

4.5.2 Results of the Clustering of Synthetic Galaxies Selected on Stellar Mass

The spatial correlation functions for all stellar masses and redshifts were reliably modeled as power laws, with $\gamma = 1.8$, over the range of scales probed. Figure 4.8 shows the spatial correlation functions for each stellar mass-redshift catalogue.

Figure 4.9 shows how the comoving clustering strength of the galaxies varied with stellar mass. Comparisons between samples of different mass should only be made at the same redshift otherwise changes in r_0 due to evolution with redshift complicate the picture. Galaxies of mass $\log(M_{\star}/\mathrm{M}_{\odot})$=9.3, 9.7 and 10.2 at $z = 0.09$ had the same clustering strength. This implies such galaxies are typically found in the same mass haloes or at least the same range of halo masses. At all other redshifts more massive galaxies were more clustered with the increase in clustering strength with stellar mass being more pronounced at higher redshifts. More massive galaxies typically reside in more massive haloes and hence cluster more strongly.

Figure 4.10(a) shows the comoving spatial clustering strength, r_0, plotted against red-

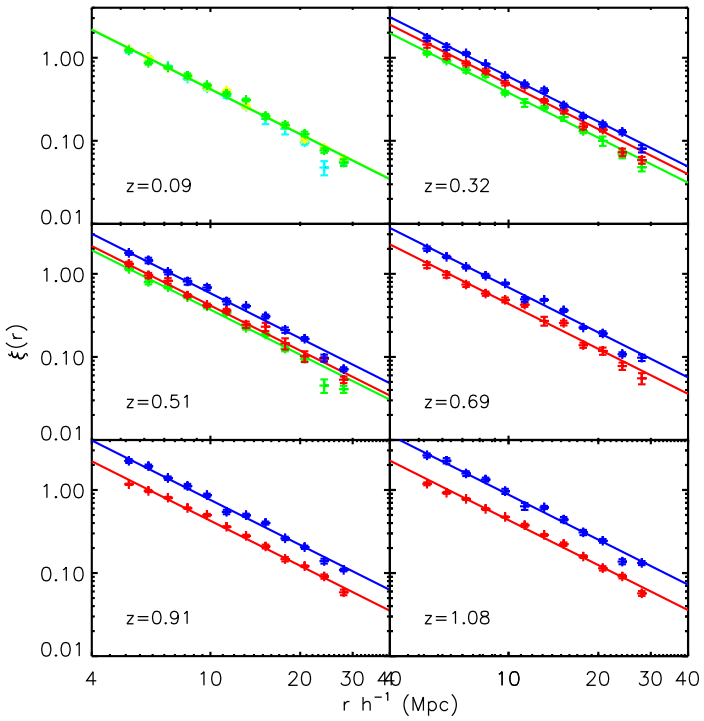


Figure 4.8: The spatial correlation function for galaxies selected on stellar mass and redshift from the Delucia et. al. 2007 semi-analytic model. Each panel shows a different redshift. Correlation functions for galaxies of mass $\log(M_\star/\mathrm{M}_\odot)$=9.3, 9.7, 10.2, 10.8, 11.2 are plotted as cyan, yellow, green, red and blue, respectively.

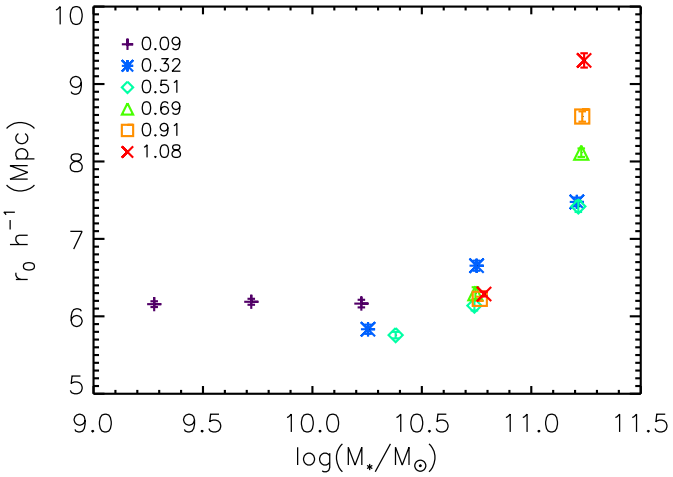
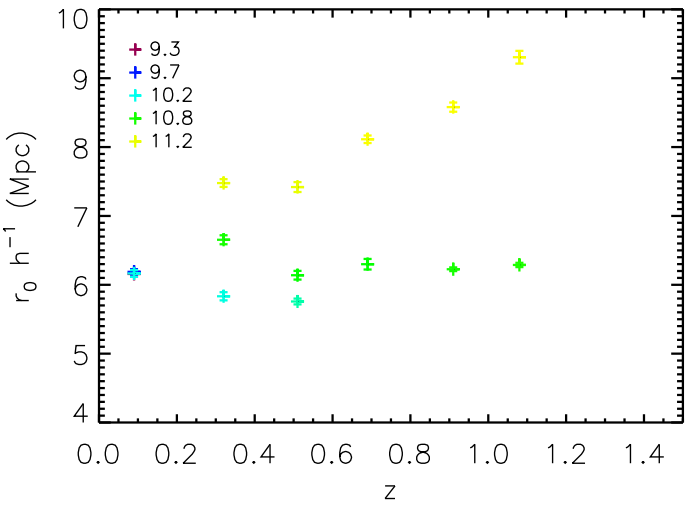


Figure 4.9: The comoving spatial clustering strength as a function of stellar mass for galaxies from the semi-analytic model of Delucia et. al. 2007. The redshift of each galaxy sample is shown in the legend.

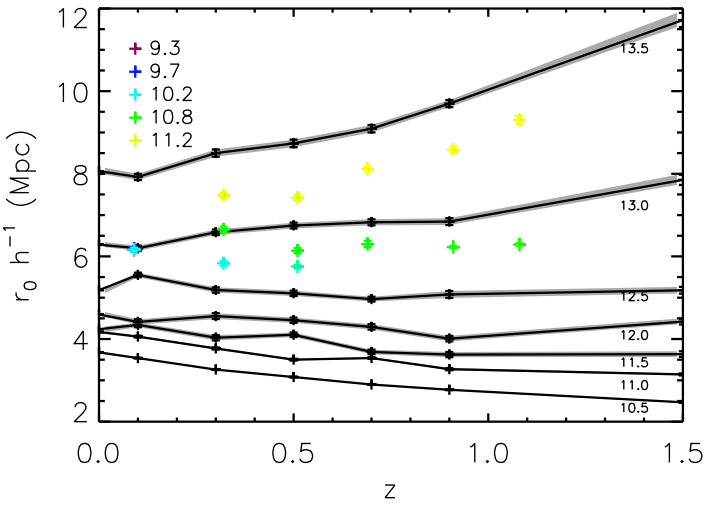
shift, for each sample. Galaxies of stellar mass $\log(M_\star/\mathrm{M}_\odot) = 10.2$ showed a decrease in clustering strength over $0.09 < z < 0.51$. The clustering strength of galaxies with stellar mass $\log(M_\star/\mathrm{M}_\odot) = 10.8$ showed a decrease between $z = 0.32$ and $z = 0.51$ but remained constant over $0.51 < z < 1.08$. The comoving clustering strength of galaxies of stellar mass $\log(M_\star/\mathrm{M}_\odot) = 11.2$ increased by $\sim 25\%$ over $0.51 < z < 1.08$.

The r_0 values seem to be systematically too high for the stellar masses considered here. We compare these values to those of Meneux et al.(2008) which were selected on stellar mass in the VVDS. Meneux et al.(2008) found $r_0 = 4.28^{+0.43}_{-0.45}$ h^{-1}Mpc for $\log(M_\star/\mathrm{M}_\odot) > 10.5$ galaxies at $z = 0.85$. The synthetic galaxies of the same mass at this redshift had $r_0 \sim 7$ h^{-1}Mpc, i.e. the clustering strength was $\sim 60\%$ larger. Ichikawa et al.(2007) measured the clustering of $\log(M_\star/\mathrm{M}_\odot) = 10.6$ galaxies at $z = 1.4$ where they found $r_0 = 4.6^{+1.6}_{-2.5}$ h^{-1}Mpc. These very large errors means their value was just about comparable with the synthetic galaxies, if we extrapolated their clustering strengths to $z = 1$ and taking into account the slight mass difference.

We attribute the high clustering strengths in the model to the fact that the simulations are too reliant on sustained star formation, whereas real galaxies form stars through lots of bursts. Hence the stellar mass of the simulated galaxies increases a lot (almost doubles) between $z \sim 1$ and $z = 0$ as sustained star formation continues. Therefore massive galaxies



(a) Galaxies.



(b) Galaxies and haloes.

Figure 4.10: The comoving spatial clustering strength of galaxies selected on stellar mass as a function of redshift from the semi-analytic model of Delucia et. al. 2007 over $0 < z < 1.5$. Panel a) shows just the galaxies. Panel b) also shows the dark matter halo clustering strengths for haloes of fixed mass. The log of the stellar mass in units of solar mass of each galaxy sample is shown in the legend in the top left. The halo masses are indicated on the right hand side of the plot. Note the data points at $z = 0.09$ for $\log(M_\star/\mathrm{M}_\odot) = 9.3, 9.7$ are hidden under the data point for $\log(M_\star/\mathrm{M}_\odot) = 10.2$ galaxies.

in the simulations are too rare at $0.5 \leq z \leq 1$ and hence only exist in massive haloes which leads to their amplified clustering strengths.

Figure 4.10(b) shows the comoving spatial clustering strengths for the synthetic galaxies, however this time the clustering strengths of haloes of fixed mass are also plotted. The variation in clustering strength with redshift for $\log(M_\star/\mathrm{M}_\odot) = 11.2$ galaxies was similar to that for high mass haloes ($\log(M_{hz}/\mathrm{M}_\odot) \sim 13.5$) which showed an increase in clustering strength with redshift. This implies that $\log(M_\star/\mathrm{M}_\odot) = 11.2$ galaxies had a constant biasing, with respect to the dark matter haloes, with redshift. Figure 4.10(b) also indicates galaxies with $\log(M_\star/\mathrm{M}_\odot) = 10.8$ were found in $\log(M_{hz}/\mathrm{M}_\odot) \sim 12.9$ haloes which showed a constant clustering strength with redshift.

4.6 Summary

In this chapter the clustering of dark matter haloes from the Millennium Simulation was measured on large scales for a range of halo masses and redshifts. The clustering of haloes of the same mass at different redshifts increased for high mass haloes ($\log(M_{hz}/\mathrm{M}_\odot) \geq 12.5$) and decreased slightly for low mass haloes ($\log(M_{hz}/\mathrm{M}_\odot) \leq 12.0$). We attributed this trend to rarer haloes being more clustered as they tend to be found close to one another. We are able to rule out formation redshift as a dominant process in the clustering of haloes as studies in the literature indicate it should produce trends opposite to what is observed here. We discussed how other halo properties such as halo concentration and halo spin scale with environment. The clustering of synthetic galaxies from the semi-analytic model of De Lucia & Blaizot(2007) was also measured on large scales. Apart from at $z = 0.09$, more massive galaxies were found to be more clustered. The clustering strength of $\log(M_\star/\mathrm{M}_\odot) = 10.2$ galaxies decreased over $0.1 < z < 0.5$, whereas $\log(M_\star/\mathrm{M}_\odot) = 10.8$ galaxies remained constant and $\log(M_\star/\mathrm{M}_\odot) = 11.2$ galaxies increased. These clustering variations with redshift broadly matched the clustering of the dark matter haloes implying the biasing of such galaxies is constant with redshift. The synthetic galaxies seem too clustered when compared with values from observational data which we conclude is because the simulations are too reliant on sustained star formation.

Chapter 5

The Clustering of SWIRE Galaxies

5.1 Introduction

Previous clustering analyses, particularly at high redshift, have often been restricted to galaxy samples which are small, selected on observable properties and/or a mixture of different spectral types (e.g. Waddington et al.(2007), de la Torre et al.(2007), Ilbert et al.(2006), Pollo et al.(2006), Le Fèvre et al.(2005)). Small samples lead to limited statistics and the inability to break the sample into interesting subsamples based on properties, e.g. by luminosity or SFR. Selecting samples based on observable properties leads to varying intrinsic properties with redshift and hence makes interpreting the evolution of clustering difficult. Mixing spectral types makes it impossible to separate the contribution to the clustering signal from early and late-type galaxies which tend to cluster very differently.

This thesis hopes to overcome some of the limitations of earlier clustering analyses by using the SWIRE photometric redshift catalogue of Rowan-Robinson et al.(2008) to measure the clustering of galaxies. The complicated selection effects inherent to the photometric redshift catalogue were thoroughly investigated in Chapter 3 and are constrained in this chapter so the potential scientific benefits of the catalogue can be exploited. The SWIRE catalogues covered multiple fields and large volumes providing a large sample of galaxies. The photometric redshifts enabled the evolution of the galaxy clustering to be investigated for a much larger dataset than would be feasible with spectroscopic redshifts.

The template fitting used in Rowan-Robinson et al.(2008) determined photometric redshifts with an rms accuracy at the 4% level. The template fitting procedure also provided spectral classifications as well as intrinsic properties such as stellar mass estimates and $SFRs$ thereby enabling the clustering of early and late-type galaxies to be analysed as a function of these properties. The clustering results are presented and are discussed in the context of current theories of galaxy evolution.

5.2 Measuring the Clustering of SWIRE Galaxies

The clustering analyses presented here utilized the photometric redshift catalogues in the ELAIS-N1 and ELAIS-N2 fields covering 10.34 deg^2. The other four SWIRE fields were not used for the following reasons. CDFS and Lockman hole only had 3 band optical photometry (g,r,i) meaning the quality of the photometric redshifts was worse. ELAIS-S1 and XMM had optical data available from the ESO WFI surveys (Berta et al.(2006), Berta et al.(2008)) and CFTHLS respectively, however it is known (private communication with Rowan-Robinson) that the photometric redshifts in these fields are not as good, presumably due to the different optical data and insufficient photometric redshift calibration.

Two clustering analyses were conducted:-

- The clustering of elliptical (template E), spiral (templates Sab - Scd) and starburst (templates Sdm - sb) galaxies as a function of stellar mass and redshift.
- The clustering of spiral and starburst galaxies (Sab-sb) collectively, as a function of SFR and redshift.

We binned the galaxies into stellar mass-redshift cells $(M_\star, z)$ and SFR-redshift cells (SFR, z). The angular correlation function of the galaxies in each cell was measured. Limber's equation was used to calculate the spatial clustering strength r_0 from the amplitude of the correlation function. It should be noted some preliminary analyses were conducted, whilst the method to measure the clustering was being perfected. The results from these analyses are incorrect but they do highlight the necessity of using the full method. The preliminary analyses are discussed briefly in Section 5.7.

5.2.1 Constraining the Selection Effects

In it's raw form the photometric redshift catalogue is not suitable for a clustering analysis. It must be constrained to ensure there are no artificial variations in the spatial distribution of the galaxies resulting from the selection effects. Otherwise it would be impossible to distinguish the *real* clustering signal resulting from the large scale distribution of galaxies from that imposed by the selection effects.

For the $M_\star - z$ (or $SFR - z$) analysis the catalogues for ELAIS-N1 and ELAIS-N2 were cut at $S_{3.6\mu\mathrm{m}} > 10\mu$Jy ($S_{24\mu\mathrm{m}} > 400\mu$Jy) and $r \leq 23.2$ and the 3.6μm (24μm) mask was applied, as defined in Section 3.7. Applying the masks to the catalogues removed areas of the field with low coverage. Combined with the $S_{3.6\mu\mathrm{m}}$ ($S_{24\mu\mathrm{m}}$) cut, this ensured a high completeness of 90% and a variation in completeness of $< 2\%$. The masks also rejected areas of the field with low r band "depth". Combined with the $r \leq 23.2$ cut, this gave an r band completeness of 95% and a variation in completeness of a couple of percent. The mask also removed foreground stars which obscure galaxies along their line of sight.

Applying these criteria ensured the data used was homogeneous and still retained a sufficiently large number of galaxies. Before the n_{band} and χ^2 criteria, designed to select good photometric redshifts, can be applied the $(M_\star, z)$ and (SFR, z) cells need to be properly defined.

5.2.2 Defining Stellar Mass - Redshift Cells

We binned the galaxies into redshift slices of $\Delta z = 0.2$ over $0.1 < z < 1.6$ and by stellar mass in bins of approximately $d\log(M_\star/\mathrm{M}_\odot) = 0.5$. The bins were defined to give samples of constant mass over the range of redshifts and a range of masses at each redshift (where possible). The galaxies in each $(M_\star, z)$ cell were split into elliptical, spiral and starburst galaxies based on the template fits of Rowan-Robinson et al.(2008), in order to analyse the clustering of the three spectral types as a function of stellar mass and redshift. We defined 21 $(M_\star, z)$ cells, see Tables 5.4, 5.6 and 5.8 for information on the number of elliptical, spiral and starburst galaxies, respectively, in each cell and for the mean stellar masses and redshifts. Figure 5.1 shows how the three types of galaxies are distributed in the stellar mass-redshift plane and the green dot-dash lines indicate the cell boundaries. The cells were defined so that the elliptical, spiral and starburst galaxies in each cell had approximately the same mean stellar mass and redshift so comparisons between the

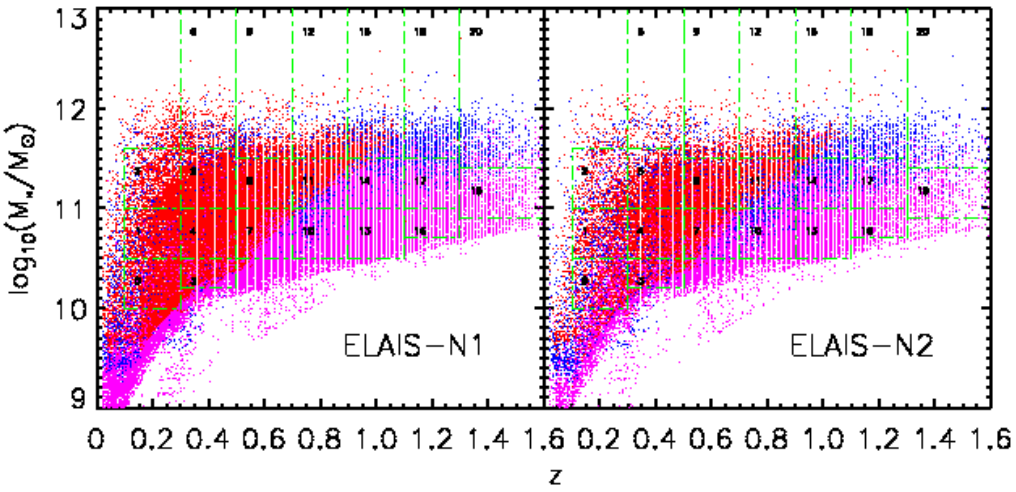


Figure 5.1: Galaxies with $S_{3.6\mu m} > 10\mu$Jy and $r \leq 23.2$ in the stellar mass-redshift plane. ELAIS-N1 is shown in the left hand panel and ELAIS-N2 in the right hand panel. Elliptical galaxies are plotted in red, spiral galaxies in blue and starburst galaxies in pink. The numbered green square cells define the stellar mass and redshift limits used in the ($M_\star$, z) analysis. See Tables 5.4, 5.6, 5.8 for more details about elliptical, spiral and starburst galaxies in each cell.

clustering of the different types at fixed stellar mass could be made.

We measured the clustering of elliptical, spiral and starburst galaxies separately. This was done to ensure any stellar mass incompleteness or change in the relative mix of spectral types due to r_{cut} (see Section 3.8) was restricted to a few $(M_\star, z)$ cells. Such cells were identified by inspection of the stellar mass-redshift plane to see which cells where intersected by the r_{cut} stellar mass limit. These cells were then removed. For elliptical galaxies cells 3, 7, 11 and 13 were removed. For spiral galaxies cells 3, 10, 13, 16 and 19 were removed.

After applying the cuts and the mask and removing cells affected by r_{cut}, 10 of the 21 cells contained a sufficient number of galaxies ($\gtrapprox 500$) in which to measure the clustering of elliptical galaxies probing $0.1 < z < 1.0$ and $10.5 < \log(M_\star/\mathrm{M}_\odot) < 11.5$. 14 cells had enough spiral galaxies probing $0.1 < z < 1.6$ and $10.0 < \log(M_\star/\mathrm{M}_\odot) < 12.0$ and there were 12 cells for starburst galaxies over $0.1 < z < 1.6$ and $10.0 < \log(M_\star/\mathrm{M}_\odot) < 11.5$. The same cells were used for each field to make cross-field comparisons easier.

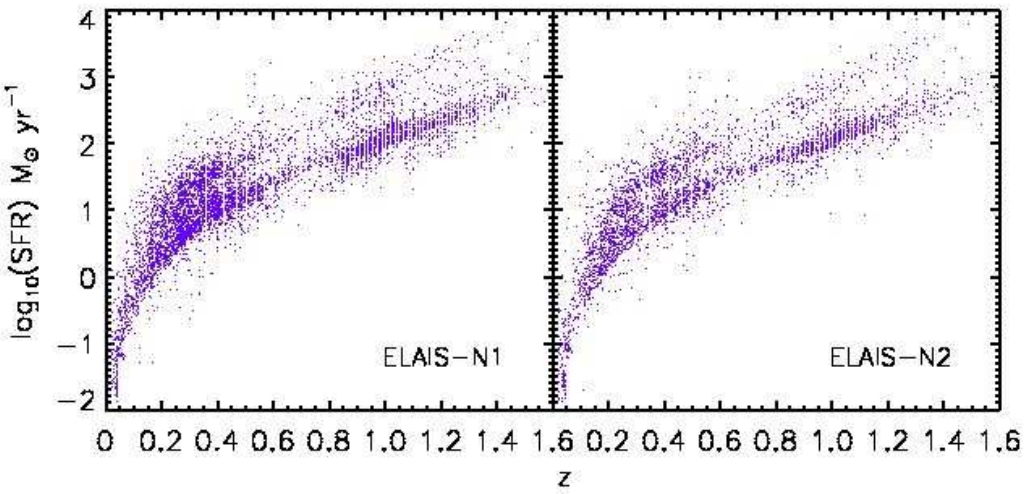


Figure 5.2: Spiral galaxies (templates Sab-sb) with $S_{24\mu m} > 400\mu$Jy and $r \leq 23.2$ in the SFR-redshift plane. ELAIS-N1 is shown in the left hand panel and ELAIS-N2 in the right hand panel.

5.2.3 Defining SFR - Redshift Cells

We binned the galaxies into 4 redshift bins in order to analyse the variation in clustering over time. The lowest redshift bin at $z = 0.3$ contained a large number of galaxies (~ 2300) which allowed us to define two SFR bins with mean SFRs of $\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) = 0.5$ and $\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) = 1.3$. The mean SFR of the three other bins increased with redshift and could not be constrained due to having fewer galaxies with SFR estimates, see Table 5.10 for more details of each cell. Therefore the mean of every bin was different. Due to the small number of galaxies it was not possible to analyse the clustering of different galaxy types. The 389 elliptical galaxies (across all redshifts) with SFR estimates were removed so the clustering of just spiral galaxies (Sab - sb) could be analysed as a function of redshift. Figure 5.2 shows the spiral galaxies in the SFR-redshift plane. We used the same cells for each field to make cross-field comparisons easier.

5.2.4 Modeling the Redshift Criteria

We applied the n_{band} and χ^2 cuts, defined in Section 3.7.4, to the galaxies in each cell to minimise the number of photometric redshift outliers. However these cuts removed sources from across the field which would have affected the measured clustering signal. This affect

was compensated for as follows.

Consider cell i. For the stellar mass-redshift analysis we measured the mean 3.6μm flux (or 24μm flux for the $SFR-z$ analysis) and mean r-band magnitude in cell i. For the stellar mass-redshift analysis we did this for ellipticals, spirals and starburst galaxies separately. All galaxies across the whole field (across all redshifts) within 1σ (i.e 1 standard deviation) of the mean flux and magnitude were selected, excluding those galaxies within the cell in question because their angular distribution is the clustering signal we are trying to measure. This selected all sources across all redshifts with similar observable properties to those in the cell. The large redshift range over which these sources were selected ensured any clustering signal present at each redshift was washed out. Such sources would respond to the n_{band} and χ^2 cuts in a similar manner to the sources in cell i due to their similar observable properties.

The field was binned into α and δ bins of 0.2×0.2 deg^2. We counted the number of galaxies within the 1σ ranges in each bin, n_{sim}. We then applied the n_{band} and χ^2 constraints for cell i to these galaxies and counted the number excluded by the cuts, n_{ex}, in each bin. The ratio n_{ex}/n_{sim} was calculated for every bin producing, for cell i, a number density map of the field showing the fraction of galaxies in each bin which were rejected. How these maps were incorporated into the method for measuring the correlation function is outlined in Section 5.2.6.

Figures 5.3, 5.4 and 5.5 show the number density maps for the $(M_\star, z)$ analysis in ELAIS-N1 for elliptical, spiral and starburst galaxies, respectively. Figures 5.6, 5.7 and 5.8 show the maps for $(M_\star, z)$ analysis in ELAIS-N2 for elliptical, spiral and starburst galaxies, respectively. Figures 5.9 and 5.10 show the maps for (SFR, z) analysis in ELAIS-N1 and ELAIS-N2, respectively.

5.2.5 Galaxy Sample Used

The final sample of galaxies used in the clustering analyses were spread across ELAIS-N1 and ELAIS-N2. Due to the large number of galaxies in the photometric redshift catalogue we were able to define strict criteria which provided a reliable homogeneous data set but still retained a sufficiently large number of galaxies with good photometric redshifts. For the $M_\star - z$ analysis, after applying the $S_{3.6\mu m}$, r_{cut}, n_{band} and χ^2 cuts and the 3.6μm binary mask, 62,907 galaxies remained in ELAIS-N1 and 35,524 in ELAIS-N2. For the $SFR-z$ analysis after applying the $S_{24\mu m}$, r_{cut}, n_{band} and χ^2 cuts and the 24μm binary

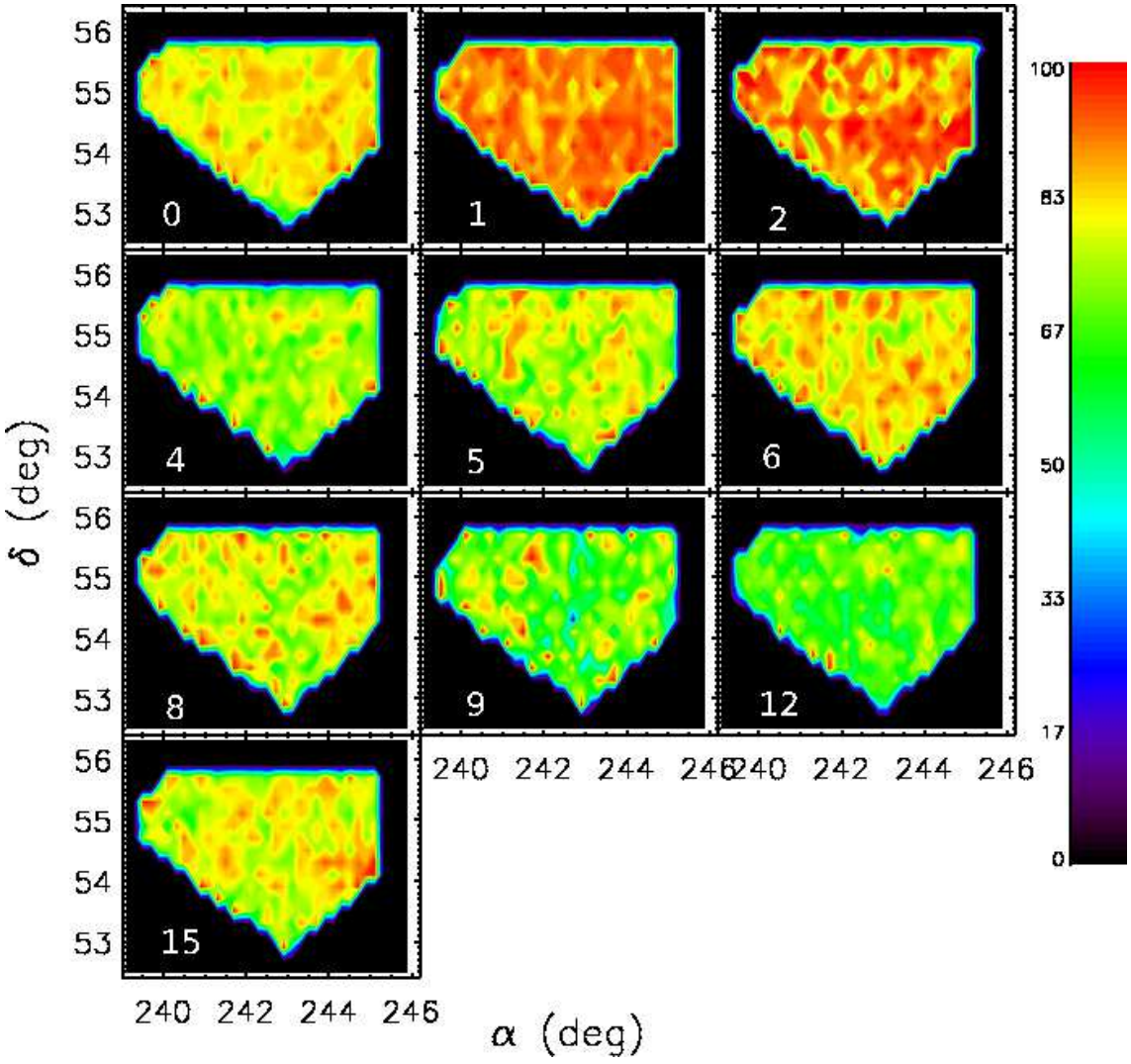


Figure 5.3: Number density maps for elliptical galaxies in each $(M_\star, z)$ cell in ELAIS-N1. The number in each panel denotes the cell number, see Table 5.4 for mass and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 3.6μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

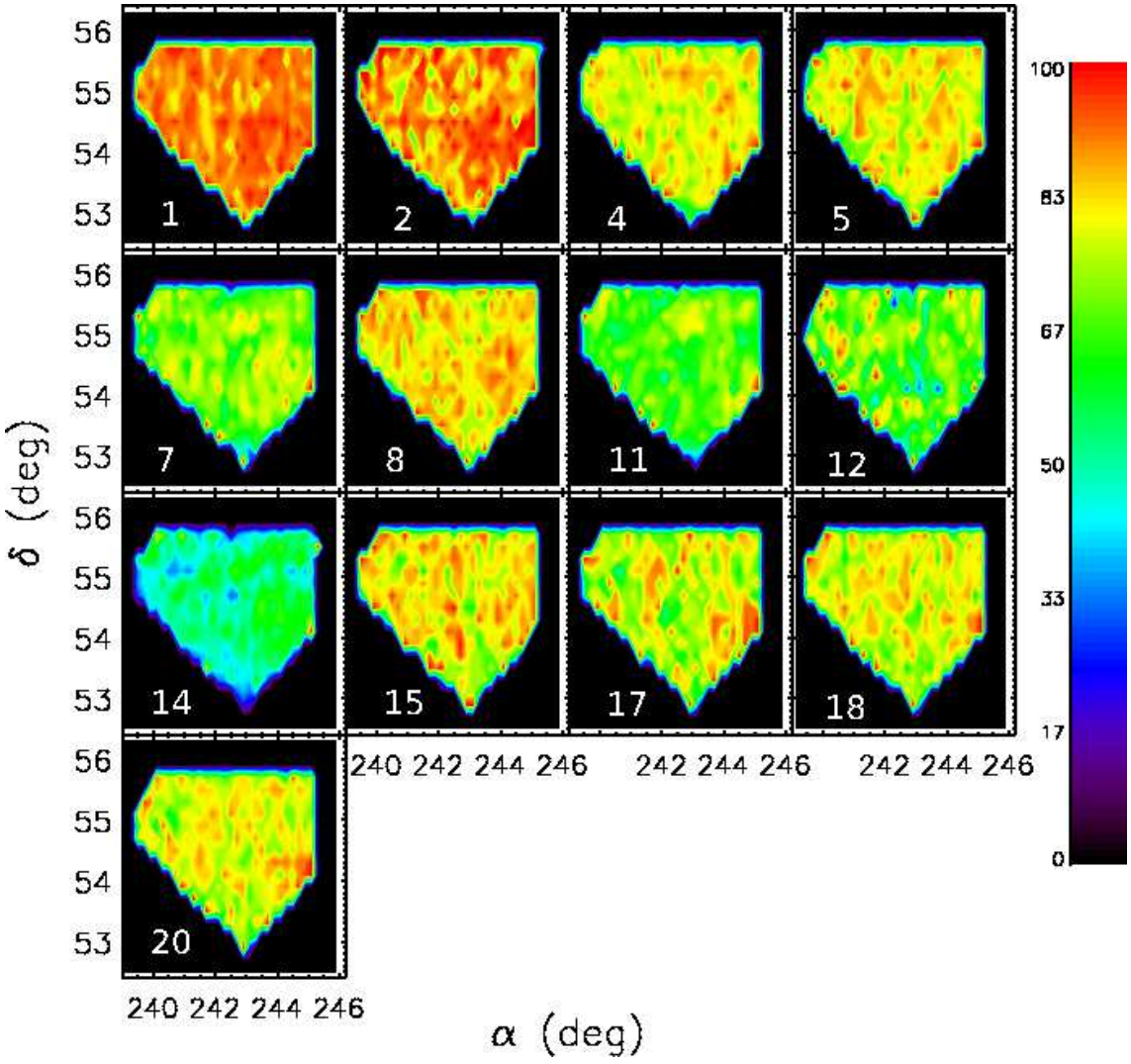


Figure 5.4: Number density maps for spiral galaxies in each $(M_\star, z)$ cell in ELAIS-N1. The number in each panel denotes the cell number, see Table 5.6 for mass and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 3.6μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

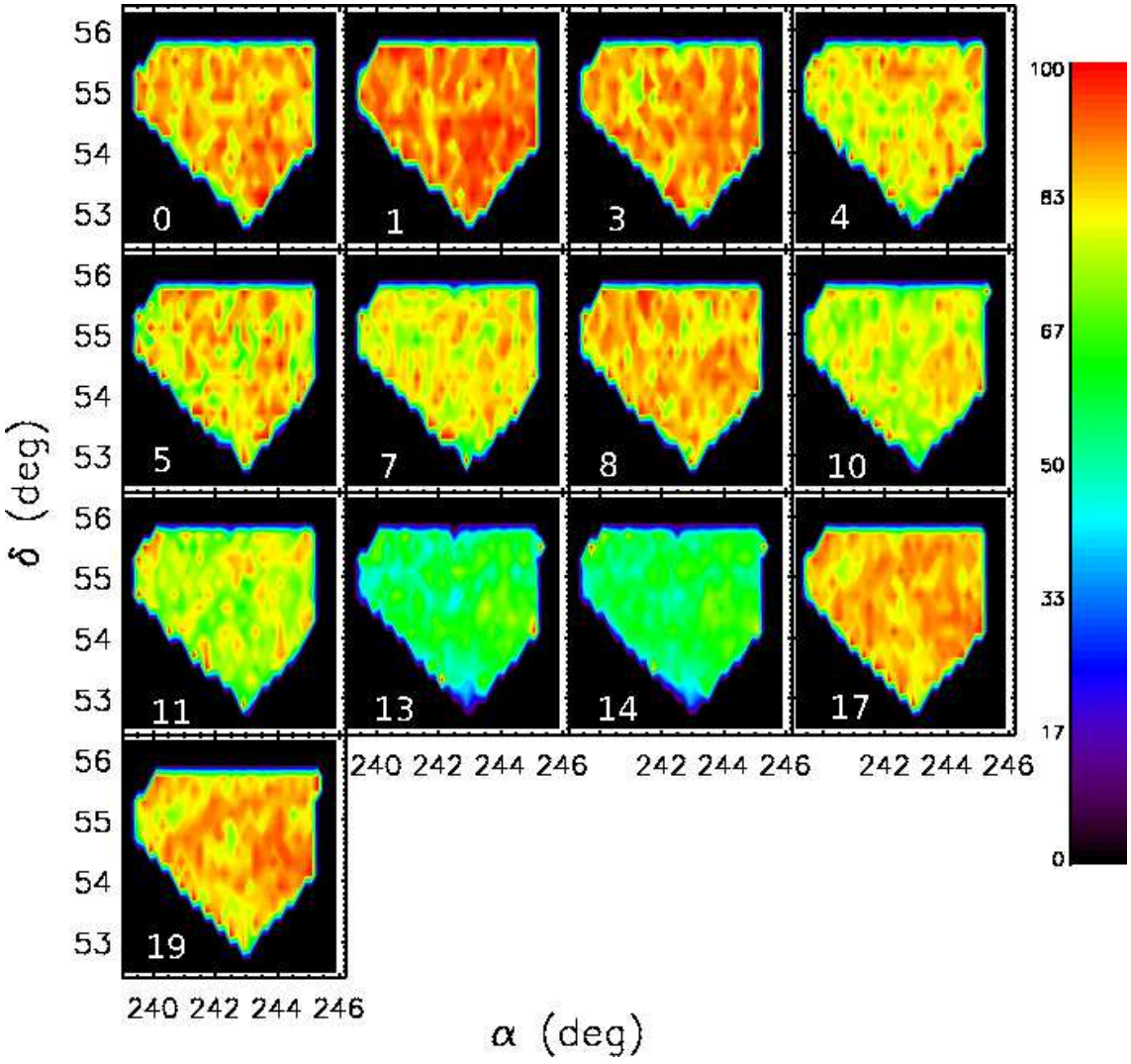


Figure 5.5: Number density maps for starburst galaxies in each $(M_\star, z)$ cell in ELAIS-N1. The number in each panel denotes the cell number, see Table 5.8 for mass and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 3.6μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

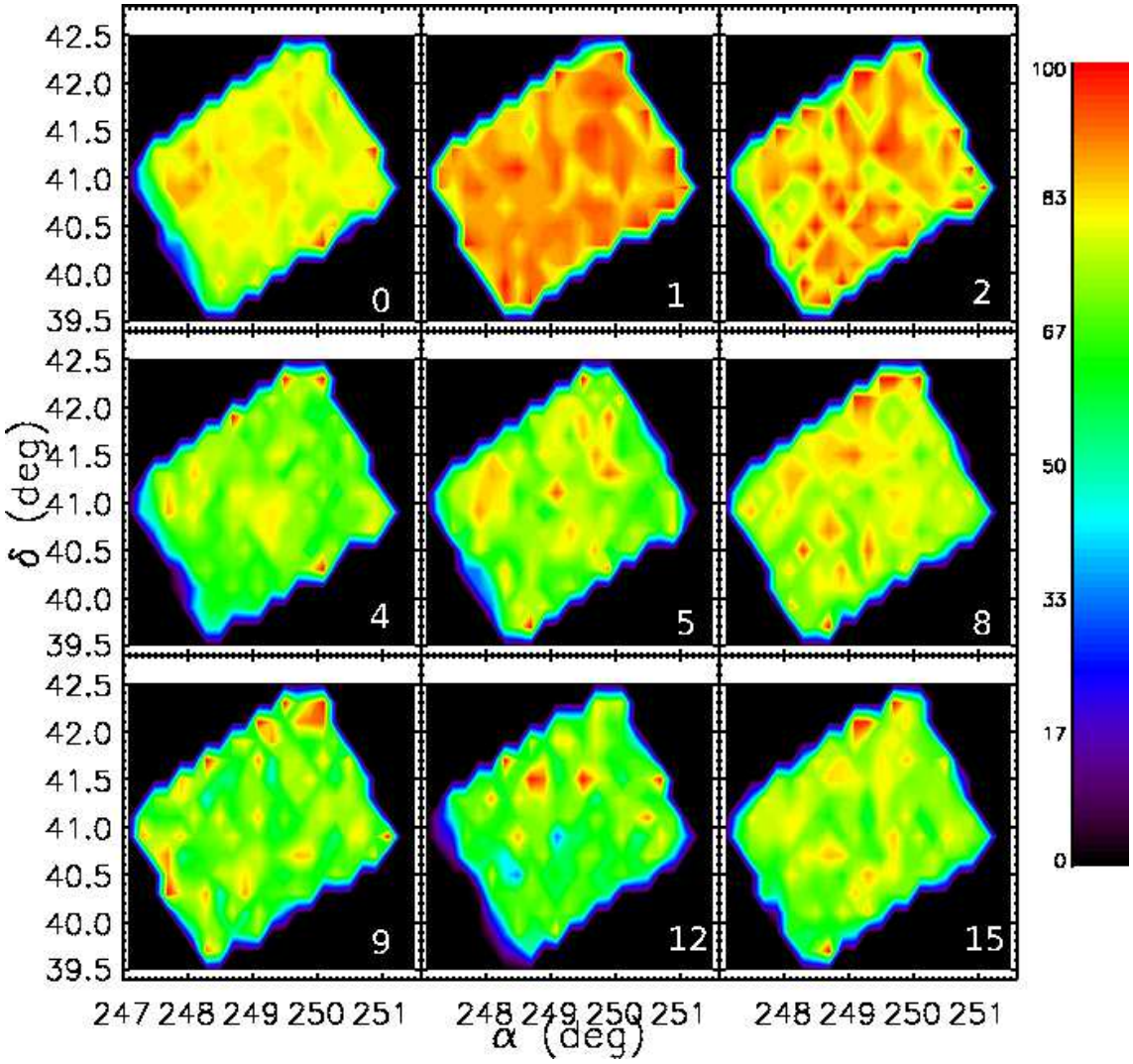


Figure 5.6: Number density maps for elliptical galaxies in each $(M_\star, z)$ cell in ELAIS-N2. The number in each panel denotes the cell number, see Table 5.4 for mass and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 3.6μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

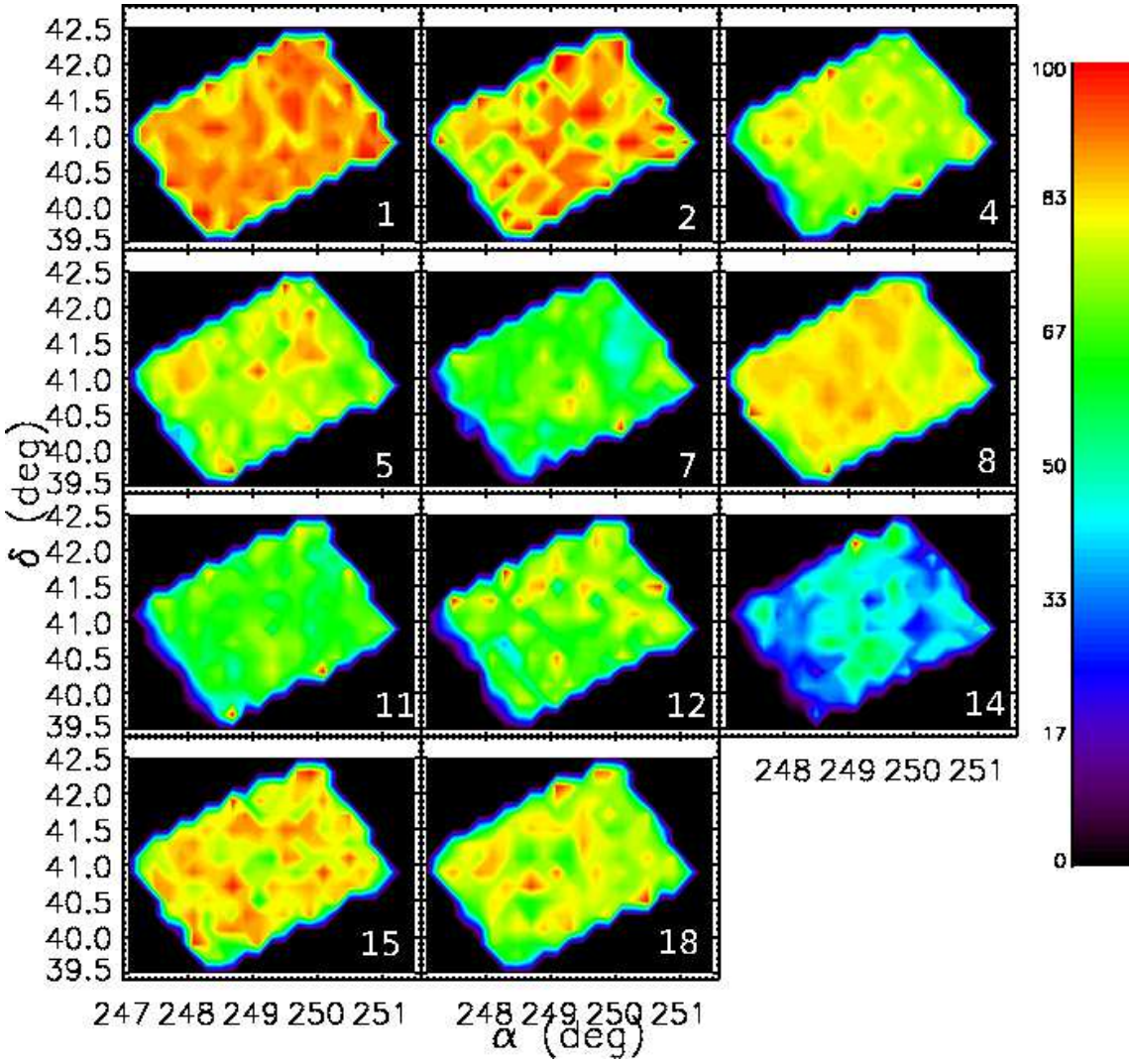


Figure 5.7: Number density maps for spiral galaxies in each $(M_\star, z)$ cell in ELAIS-N2. The number in each panel denotes the cell number, see Table 5.6 for mass and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 3.6μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

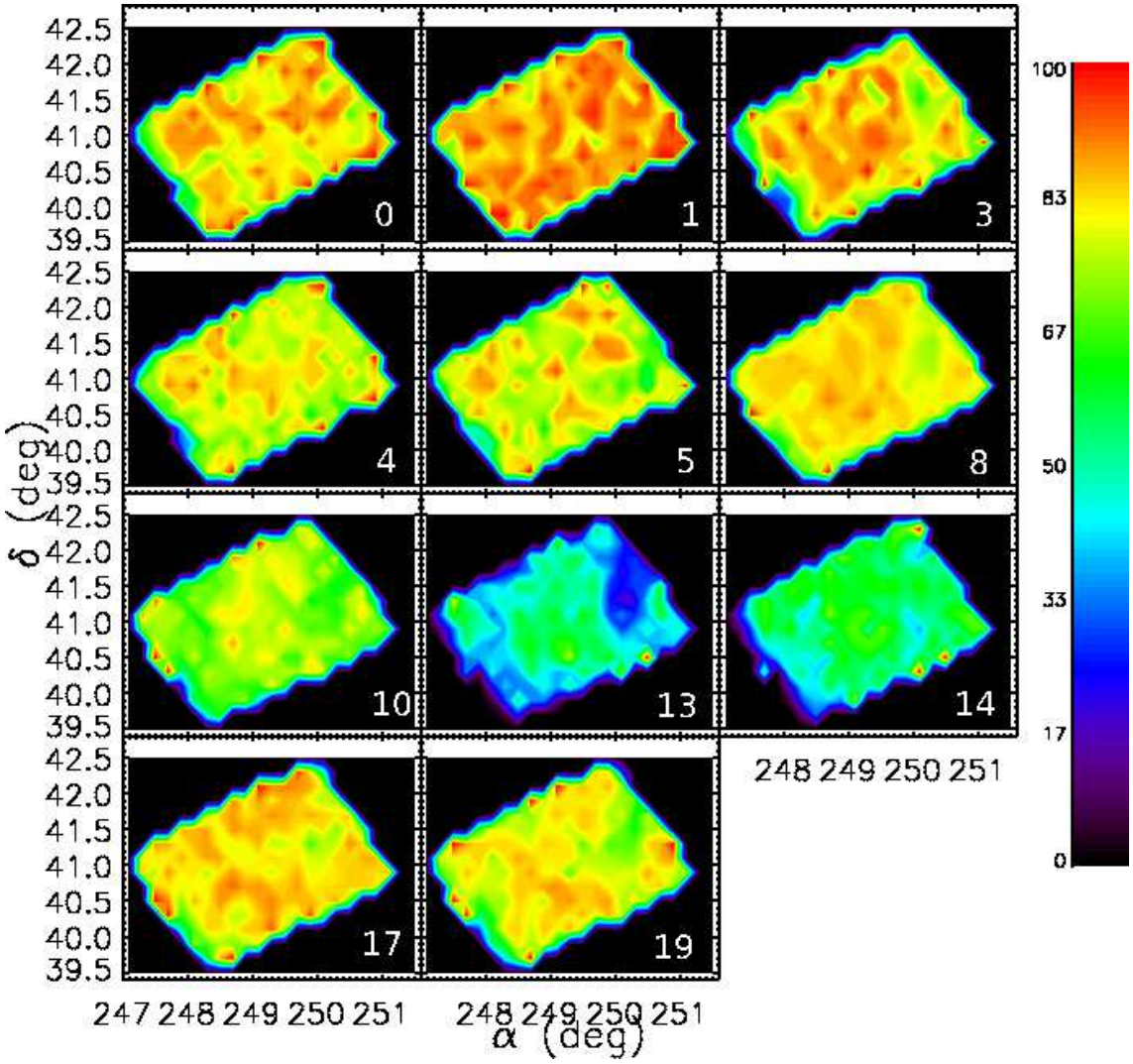


Figure 5.8: Number density maps for starburst galaxies in each $(M_\star, z)$ cell in ELAIS-N2. The number in each panel denotes the cell number, see Table 5.8 for mass and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 3.6μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

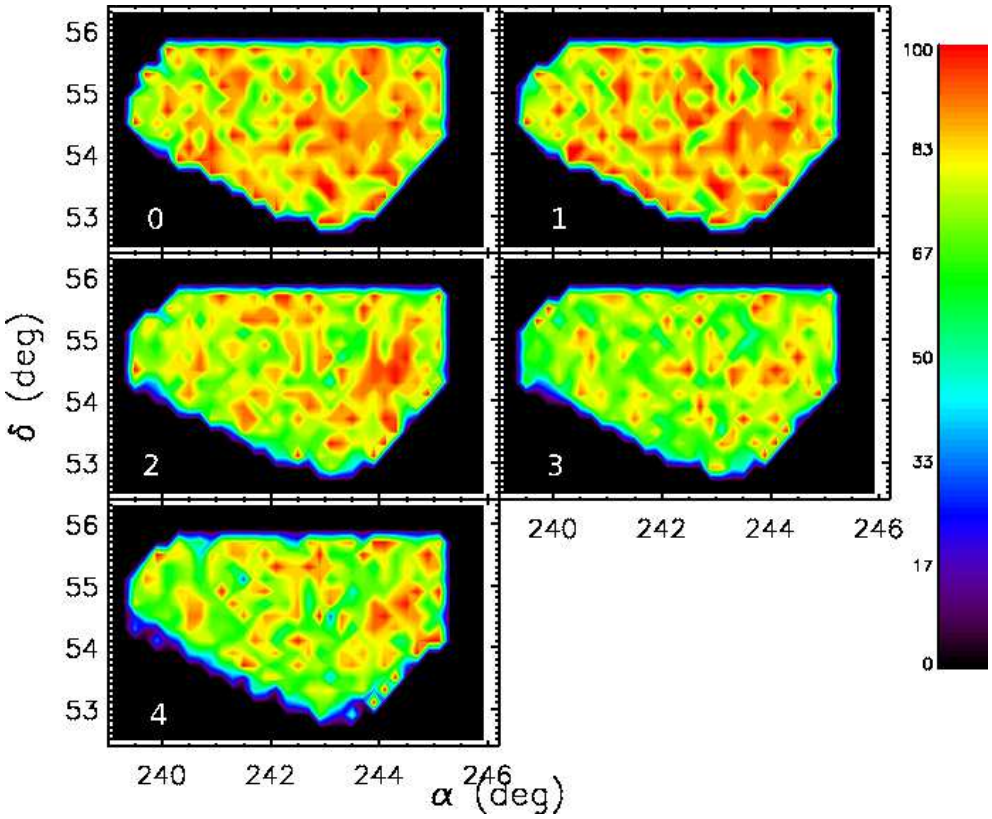


Figure 5.9: Number density maps for galaxies in each (SFR, z) cell in ELAIS-N1. The number in each panel denotes the cell number, see Table 5.10 for SFR and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 24μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

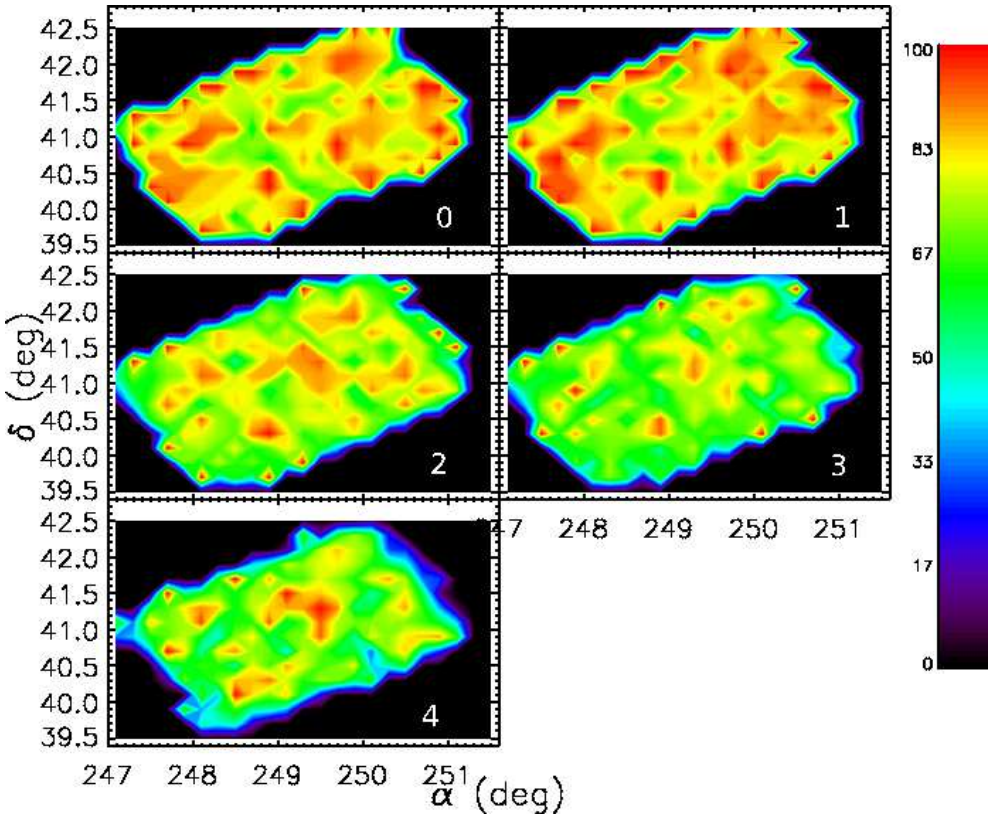


Figure 5.10: Number density maps for galaxies in each (SFR, z) cell in ELAIS-N2. The number in each panel denotes the cell number, see Table 5.10 for SFR and redshift ranges. The maps show the percentage of galaxies as a function of position with similar observable properties to those in the cell (within 1σ of the mean 24μm flux and r) which are not rejected by the n_{band} and χ^2 cuts. The blue edges are an artifact of the plotting.

Table 5.1: The SWIRE fields used in the clustering analyses along with the total area, the area of the field with optical data coverage from the INTWFS survey and the area actually used in these analyses after the coverage cuts had been applied. A large area is excluded in ELAIS-N1 where only 3 band optical photometry is available.

Field	Total Area J2000	Optical Area deg^2	Photz Area deg^2
ELAIS-N1	9.0	8.72	6.23
ELAIS-N2	4.2	4.2	4.11

mask 6296 and 3856 remained in ELAIS-N1 and ELAIS-N2, respectively. Table 5.1 shows the area of each field which was used in the analysis after applying the masks.

5.2.6 Measuring the Correlation Function

For the $M_\star - z$ analysis, three random Poisson catalogues with 500,000 sources were generated for each cell, one each for the elliptical, spiral and starburst galaxies. For the $SFR - z$ analysis only one random catalogue was required per cell, for the spiral galaxies. The random catalogues were cross-referenced with the appropriate mask (3.6μm for $M_\star - z$ and 24μm for SFR, z) and the relevant number density map to ensure they were subjected to the same constrained selection effects as the data catalogues. The number density maps were applied by binning the random catalogue in α and δ (0.2×0.2 deg^2) bins (i.e. the same binning used for the SWIRE catalogues) and then randomly removing the fraction of random sources in that bin equal to the fraction of galaxies removed in the corresponding bin on the number density map.

The data and random catalogues for each $(M_\star, z)_i$ and $(SFR, z)_i$ cell, denoted i, for a given field, f, were each split into 16 subsamples for estimating errors using a jackknife resampling technique. The field was split spatially into a 4×4 grid. Each subsample was removed in turn and the number of data-data (DD), data-random (DR) and random-random (RR) pairs separated by an angle θ were counted for the remaining subsamples using the NPT (Gray et al.(2001)) pair counting code (Section 2.6.2). The correlation function was calculated using the estimator given by Landy & Szalay(1993),

$$\omega_{ij}(\theta) = \frac{DD_{ij} - 2DR_{ij} + RR_{ij}}{RR_{ij}} \tag{5.1}$$

where j denotes the jackknife subsample and DD_j, DR_j and RR_j were all normalised by the respective total number of pairs in cell i and all a function of θ. The pairs were counted in 20 logarithmically spaced angular bins over $0.05 < \theta < 0.8$ degrees. As the galaxies in cell i were restricted to a bound area of sky we lack an estimate of the average global density for the type of galaxies in question. This was corrected for using the integral constraint which was evaluated using the random pair counts (see Section 2.4). The estimates of the correlation function for each jackknife subsample for cell i were then combined using Equation 5.2 to give the correlation function for cell i in field f, ω_{if},

$$\omega_{if}(\theta) = \sum_{j}^{N_s} \omega_{ij}(\theta)/N_s \tag{5.2}$$

where N_s was the total number of jackknife subsamples. The error on $\omega_{if}(\theta)$ was calculated using,

$$\sigma_{\omega_{if}(\theta)} = \sqrt{\frac{N_s - 1}{N_s} \frac{\sum_j^{N_s} (\omega_{ij}(\theta) - \omega_{if}(\theta))^2}{N_s}} \tag{5.3}$$

To determine the amplitude of $\omega_{if}(\theta)$, A_{if}, Equation 2.19 was fitted to ω_{if}, fixing $\gamma = 1.8$, by minimizing the χ^2 given by

$$\chi^2 = \sum_{k=1}^{n} \sum_{m=1}^{n} [\omega_t(\theta_k) - \omega_{if}(\theta_k)] H_{km} [\omega_t(\theta_m) - \omega_{if}(\theta_m)] \tag{5.4}$$

H_{km} was the inverse of the covariance matrix, σ^2_{km}, which was given by,

$$\sigma^2_{km} = \frac{N_s - 1}{N_s} \sum_{j=1}^{n} [\omega_{ij}(\theta_k) - \omega_{if}(\theta_k)][\omega_{ij}(\theta_m) - \omega_{if}(\theta_m)] \tag{5.5}$$

where N_s was the number of jackknife samples and only the covariances between adjacent bins were retained. The gradient of the correlation functions showed no variation from $\gamma = 1.8$ and hence it was fixed at this value.

The amplitude of the angular correlation function, A_{if}, was then then used to derive an estimate of the spatial clustering strength for field f, $r_{0,if}$. This was done using Limber's equation (Equation 2.21). However, in order to use Limber's equation the redshift distribution of the galaxies within the cell was required. Every galaxy, within cell i, was selected and their photometric redshift distribution was determined using the method outlined in Section 3.5.4. For each galaxy this involved turning the χ^2 distribution for the goodness of fit of the template across the whole redshift range ($0 < z < 6$) into a probability distribution. The probability distribution for each galaxy was then summed together to produce the redshift distribution. The redshift distribution was then truncated

at the redshift boundaries of the cell in question. This ignored any potential broadening of the redshift distribution from sources outside the cell which may have scattered into the cell. This also ignored any broadening produced by the galaxies within the cell. The redshift distribution and A_{if} were then fed into Limber's equation along with a value for the ϵ parameter, see Section 2.2.1.1. We used $\epsilon = \gamma - 3$ which meant $r_{0,if}$ determined from Limber's equation was effectively $r_0(z = 0)$ and was given in comoving coordinates. To minimize the effects of cosmic variance the measurements of $r_{0,if}$ from the two fields were combined using a weighted mean to give $r_{0,i}$, the spatial clustering strength for cell i

$$r_{0,i} = \frac{\sum_f^N \Omega_f r_{0,if}}{\sum_f \Omega_f} \tag{5.6}$$

where Ω_f is the area of the f^{th} field. The error on the combined $r_{0,i}$ was given by

$$\sigma_{r_{0,i}} = \sqrt{\frac{1}{N} \frac{\sum_f^N (r_{0,if} - r_{0,i}))^2}{N-1}} \tag{5.7}$$

where N is the number of fields.

5.3 Clustering of SWIRE galaxies

The following subsections present the results of the two clustering analyses. Section 5.5 discusses the results in the context of current theories of galaxy evolution and compares our findings to similar analyses in the literature.

5.3.1 Stellar Mass - Redshift Results

We measured the angular correlation function for elliptical, spiral and starburst galaxies for every $(M_\star, z)$ cell where there was a sufficient number of galaxies and where the criteria to constrain the selection effects did not produce an incompleteness in the cell. This enabled the clustering to be measured over a range of stellar masses, $10.0 < \log(M_\star/\mathrm{M}_\odot) < 12.0$ and redshifts $0.1 < z < 1.6$.

For each type of galaxy, across all stellar masses and redshifts, the correlation function was reliably modelled as a power law (Equation 2.19). Figures 5.11, 5.12 and 5.13 show the angular correlation functions in ELAIS-N1 for elliptical, spiral and starburst galaxies, respectively. Figures 5.14, 5.15 and 5.16 show the angular correlation functions in ELAIS-N2 for elliptical, spiral and starburst galaxies, respectively. The typical amplitudes for the correlation functions were $A = 0.02$. There was no increase in γ for elliptical galaxies nor

any variation with stellar mass. The comoving spatial clustering strengths, r_0, for each $(M_\star, z)$ cell are shown in Tables 5.4, 5.6 and 5.8 for elliptical, spiral and starburst galaxies, respectively.

Over the redshift range $0.1 < z < 1.0$ elliptical galaxies were the most strongly clustered with spiral and starburst galaxies being more weakly clustered at every redshift. This shows the colour-density relation was in place at $z = 1$. Starburst galaxies had similar clustering strengths to spiral galaxies except at $z = 0.2$ where their clustering strength was slightly weaker.

5.3.1.1 Variation of the Clustering Strength with Stellar Mass

The $(M_\star, z)$ cells were defined in such a way to give galaxy samples of varying stellar mass at each redshift and of constant stellar mass over a range of redshifts. The clustering strength of elliptical galaxies was invariant with stellar mass in the range $10.0 < \log(M_\star/\mathrm{M}_\odot) \leq 11.2$ ($L_B < L^*$) at all redshifts. At $z = 0.1$ $r_0 = 5.05 \pm 0.12, 4.98 \pm 0.16, 5.05 \pm 0.50$ h^{-1}Mpc for galaxies with mean stellar masses $\log(M_\star/\mathrm{M}_\odot) = 10.26, 10.77, 11.21$. Similarly at $z = 0.3$ $r_0 = 4.85 \pm 0.14, 4.77 \pm 0.11$ h^{-1}Mpc for $\log(M_\star/\mathrm{M}_\odot) = 10.78, 11.23$. At $\log(M_\star/\mathrm{M}_\odot) > 11.2$ the clustering strength increased with stellar mass content, see Figure 5.17(a). For example, for elliptical galaxies at $z = 0.4$ for $\log(M_\star/\mathrm{M}_\odot) = 11.77$ galaxies $r_0 = 6.31 \pm 0.58$ h^{-1}Mpc compared to $r_0 \sim 4.8$ h^{-1}Mpc for $\log(M_\star/\mathrm{M}_\odot) \leq 11.2$ galaxies.

Similarly, the clustering of spiral galaxies was invariant with stellar mass at $\log(M_\star/\mathrm{M}_\odot) \leq 11.2$ at all redshifts but showed a dependence on mass at $z > 0.8$ where higher stellar masses ($\log(M_\star/\mathrm{M}_\odot) = 11.6$) were probed, see Figure 5.17(b). The clustering strength of starburst galaxies showed a very weak variation with stellar mass between $\log(M_\star/\mathrm{M}_\odot) = 10.7$ and $\log(M_\star/\mathrm{M}_\odot) = 11.2$ except at $z = 0.6$ where the higher mass galaxies were much more strongly clustered, see Figure 5.17(c). The increase is unlikely to be due to a real structure at $z = 0.6$ because this would require similar structures to be present at the same redshift in both fields. The most likely explanation is that some of the starburst galaxies have been mis-classified and are actually more massive.

5.3.1.2 Relative Bias

To quantify the difference in clustering strength between elliptical and spiral galaxies, we made a simple measure of the relative bias. Following Shepherd et al. (2001), the relative

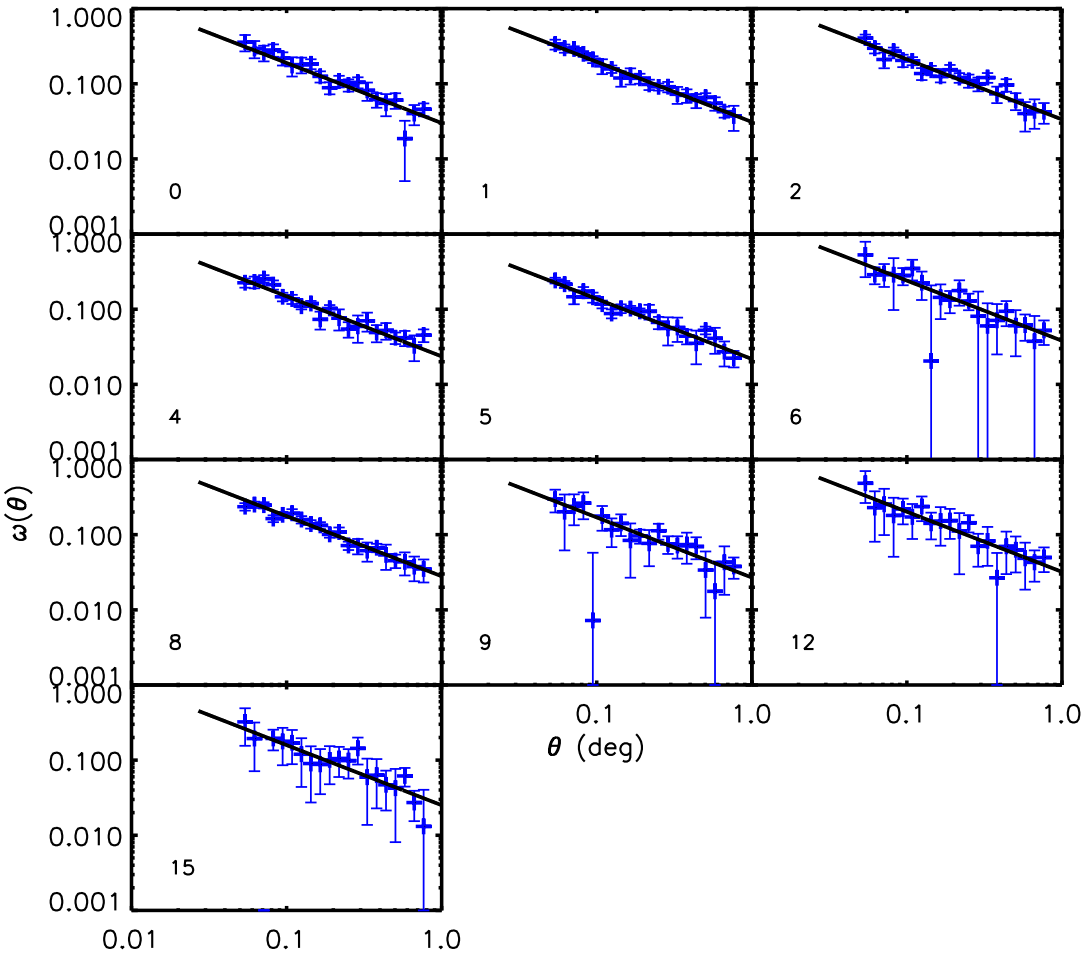


Figure 5.11: The angular correlation function, $\omega(\theta)$, for elliptical galaxies in ELAIS-N1. The power law fit was made with $\gamma = 1.8$. Each panel is for a different $(M_\star, z)$ cell, indicated by the number in the panel, see Table 5.4 for mass and redshift ranges.

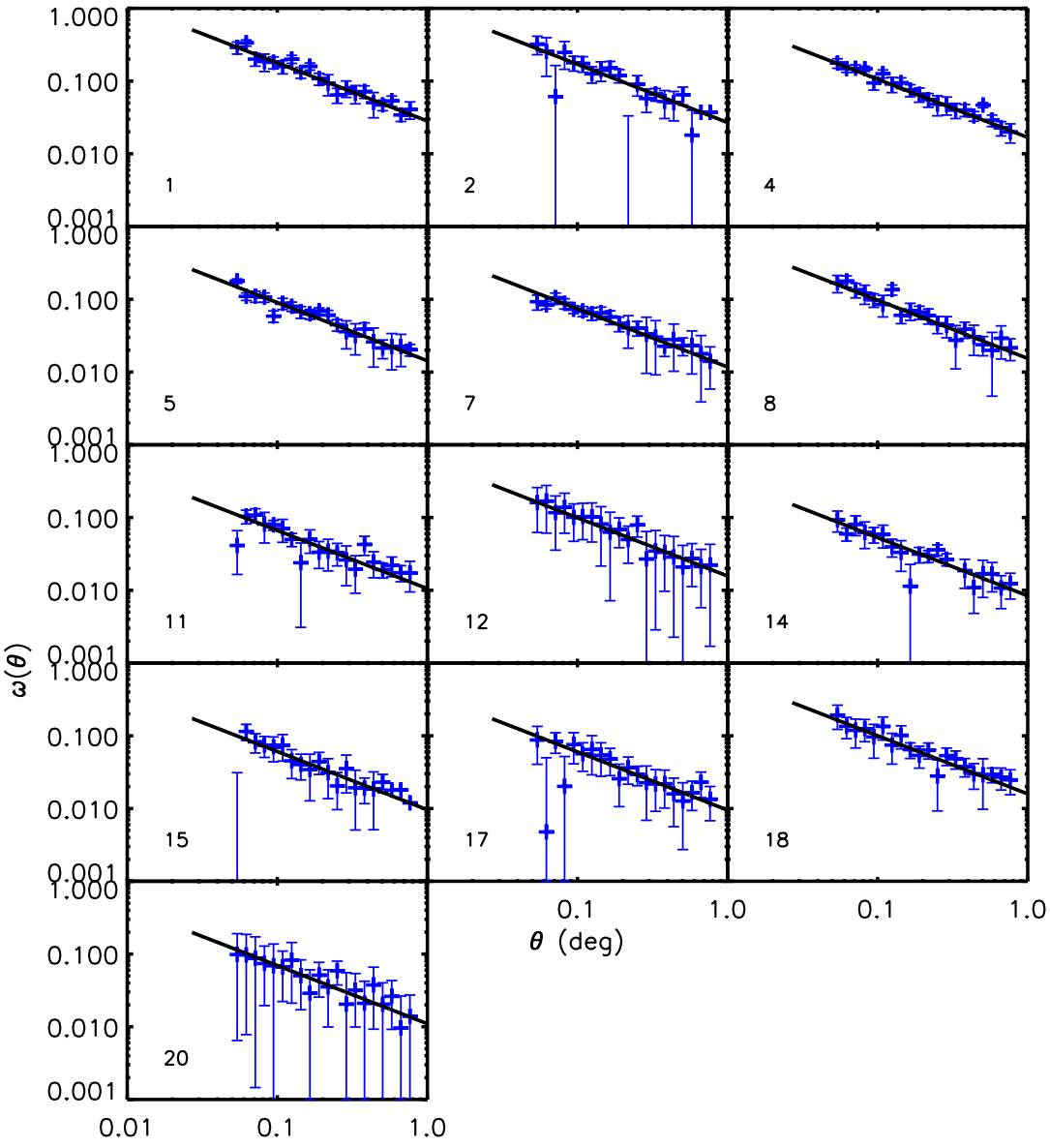


Figure 5.12: The angular correlation function, $\omega(\theta)$, for spiral galaxies in ELAIS-N1. The power law fit was made with $\gamma = 1.8$. Each panel is for a different $(M_\star, z)$ cell, indicated by the number in the panel, see Table 5.6 for mass and redshift ranges.

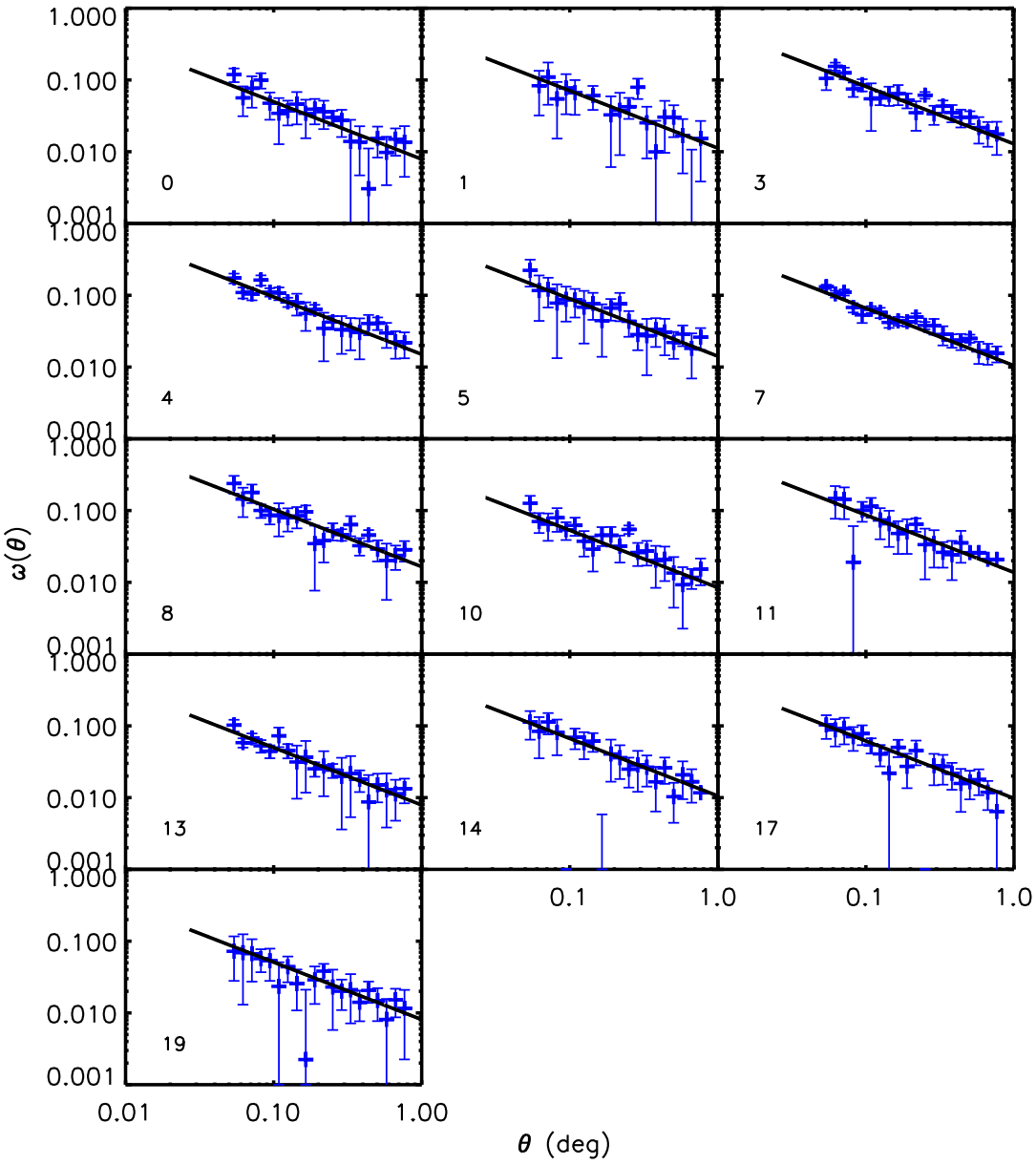


Figure 5.13: The angular correlation function, $\omega(\theta)$, for starburst galaxies in ELAIS-N1. The power law fit was made with $\gamma = 1.8$. Each panel is for a different $(M_\star, z)$ cell, indicated by the number in the panel, see Table 5.8 for mass and redshift ranges.

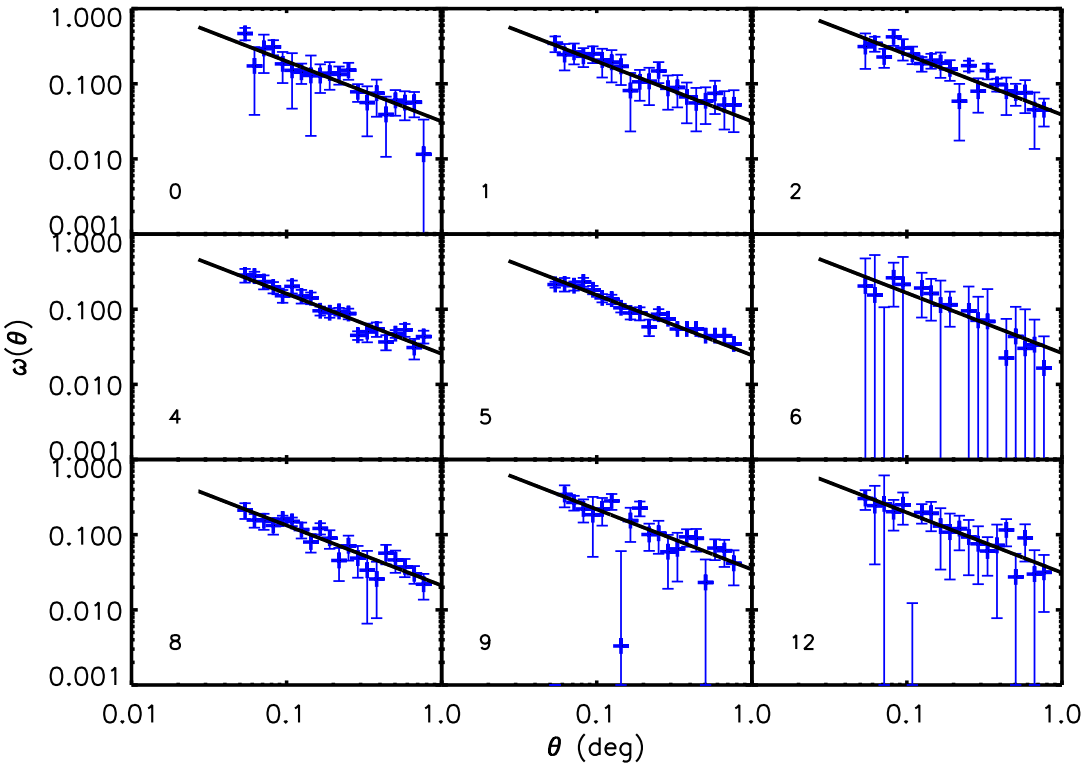


Figure 5.14: The angular correlation function, $\omega(\theta)$, for elliptical galaxies in ELAIS-N2. The power law fit was made with $\gamma = 1.8$. Each panel is for a different $(M_\star, z)$ cell, indicated by the number in the panel, see Table 5.4 for mass and redshift ranges.

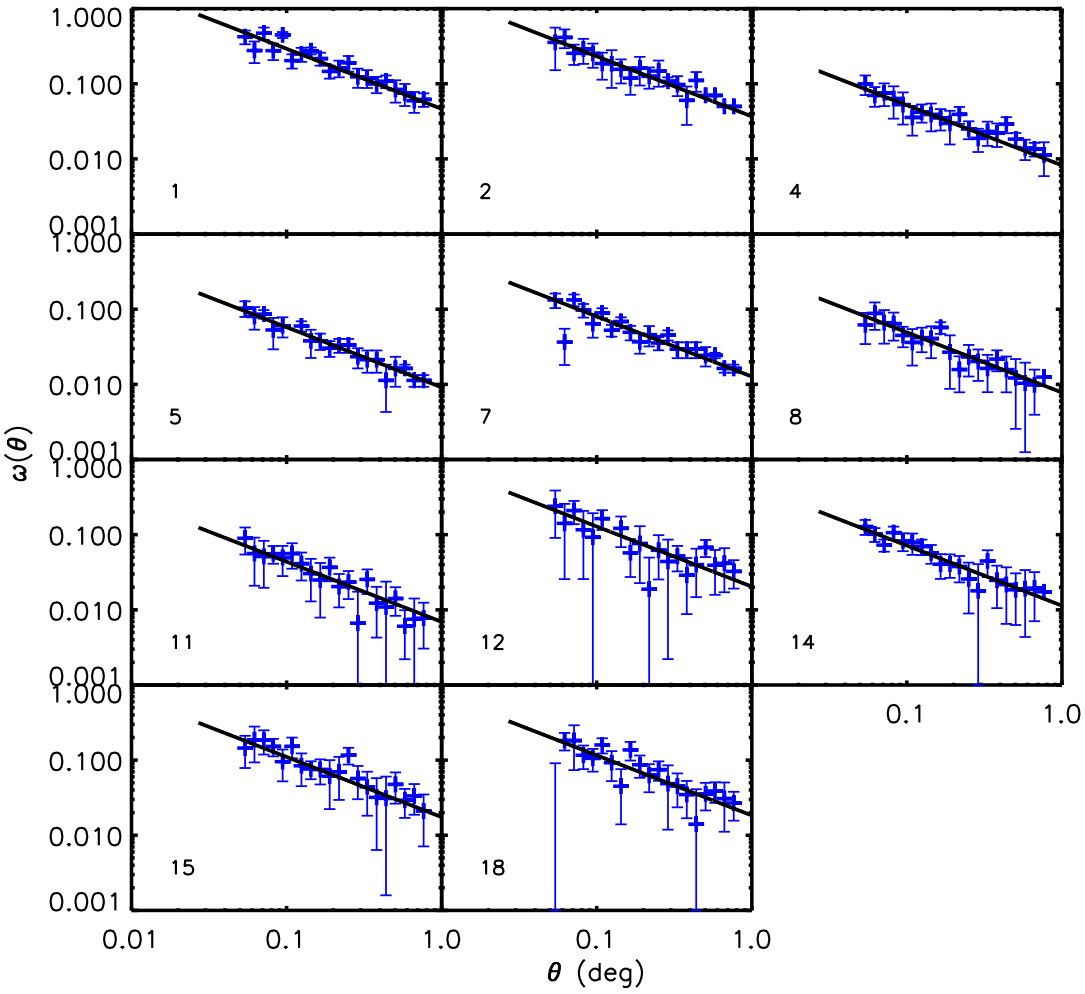


Figure 5.15: The angular correlation function, $\omega(\theta)$, for spiral galaxies in ELAIS-N2. The power law fit was made with $\gamma = 1.8$. Each panel is for a different $(M_\star, z)$ cell, indicated by the number in the panel, see Table 5.6 for mass and redshift ranges.

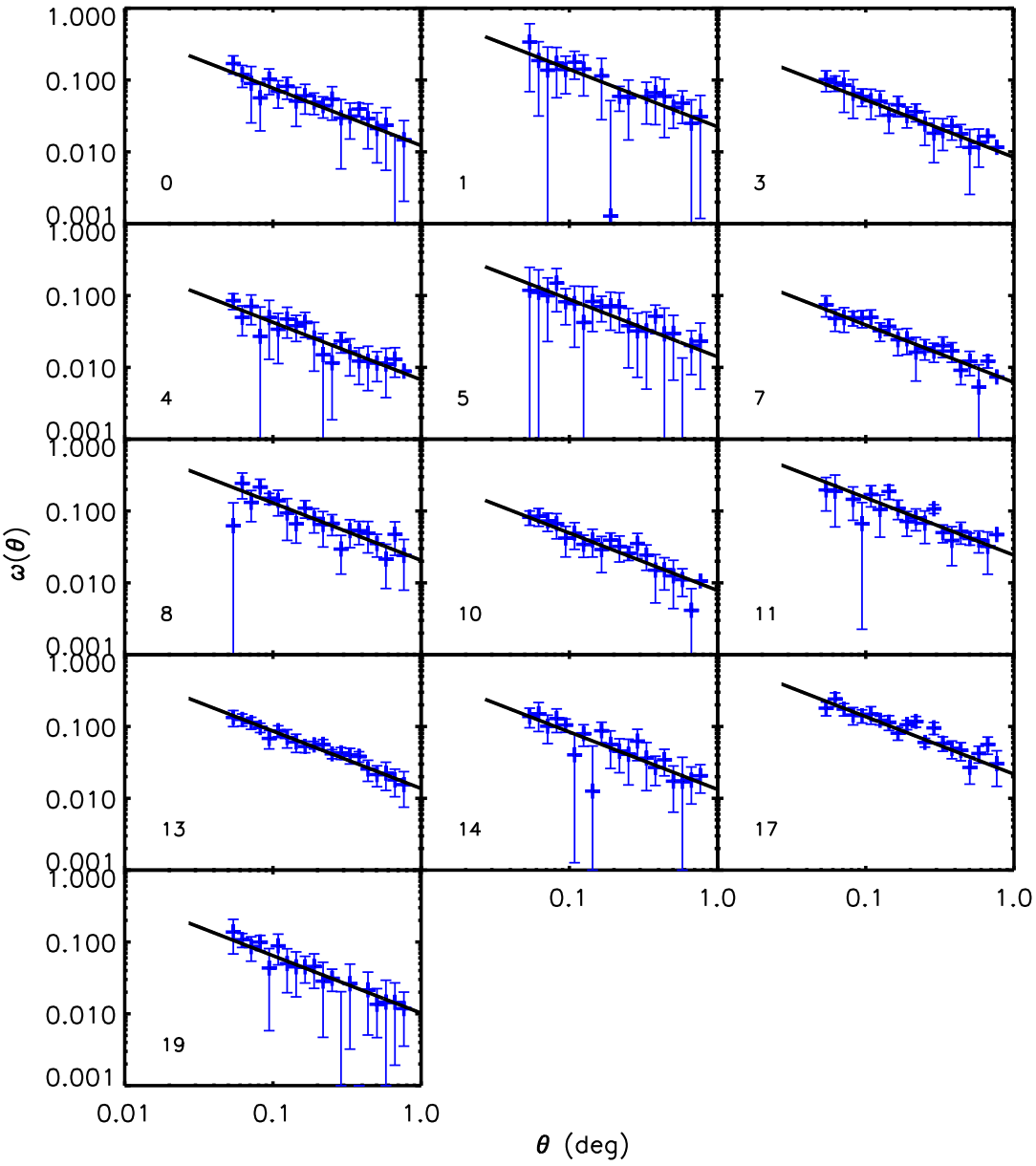
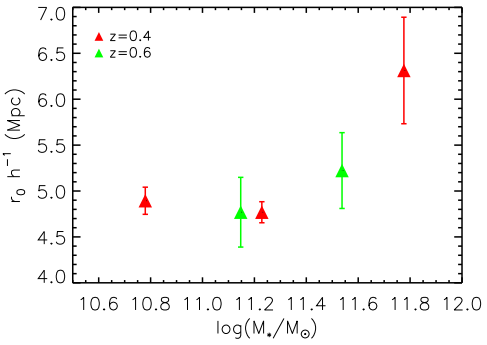
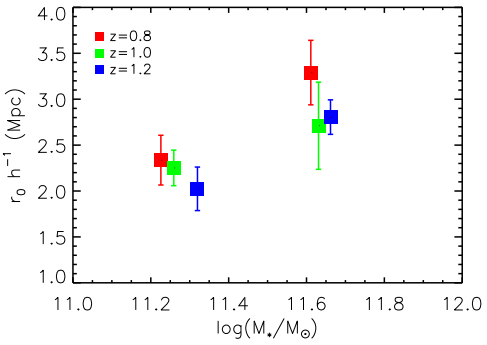


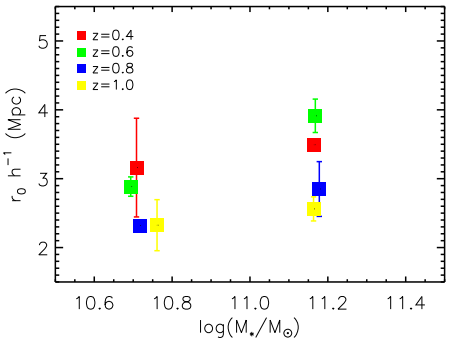
Figure 5.16: The angular correlation function, $\omega(\theta)$, for starburst galaxies in ELAIS-N2. The power law fit was made with $\gamma = 1.8$. Each panel is for a different $(M_\star, z)$ cell, indicated by the number in the panel, see Table 5.8 for mass and redshift ranges.



(a) Elliptical galaxies.



(b) Spiral galaxies.



(c) Starburst galaxies.

Figure 5.17: The comoving spatial clustering strength r_0 as a function of stellar mass in redshift bins indicated in the legends for a) elliptical galaxies, b) spiral galaxies and c) starburst galaxies.

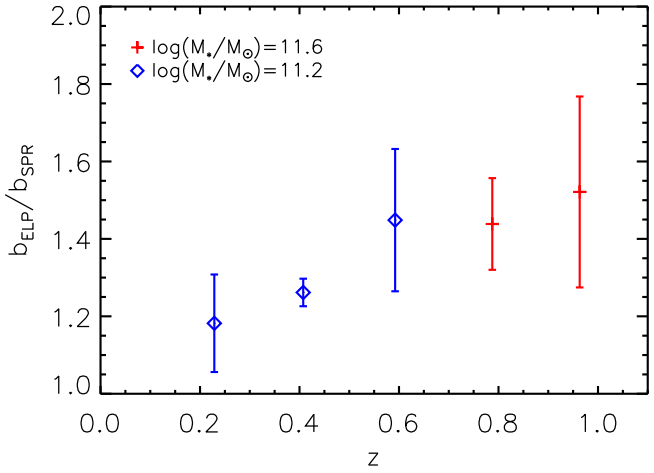


Figure 5.18: The relative bias, calculated using Equation 5.8, between elliptical and spiral galaxies with $\log(M_\star/\mathrm{M}_\odot) = 11.2$ (blue data points) and $\log(M_\star/\mathrm{M}_\odot) = 11.6$ (red data points).

bias between two types of galaxies, A and B, was defined to be

$$\frac{b_\mathrm{A}}{b_\mathrm{B}} = \sqrt{\frac{(r_0^\gamma)_\mathrm{A}}{(r_0^\gamma)_\mathrm{B}}} \tag{5.8}$$

We calculated the relative bias between elliptical and spiral galaxies for galaxies with mean stellar mass $\log(M_\star/\mathrm{M}_\odot) = 11.6$ and $\log(M_\star/\mathrm{M}_\odot) = 11.2$ as a function of redshift, see Table 5.2. The relative bias for $\log(M_\star/\mathrm{M}_\odot) = 11.2$ galaxies increased from 1.18 ± 0.13 at $z = 0.2$ to 1.45 ± 0.18 at $z = 0.6$, although could be considered to be constant to within errors. For $\log(M_\star/\mathrm{M}_\odot) = 11.6$ galaxies the relative bias was approximately constant with 1.44 ± 0.12 at $z = 0.8$ to 1.52 ± 0.25 at $z = 1.0$, see Figure 5.18.

The relative bias was also calculated between spiral and starburst galaxies with $\log(M_\star/\mathrm{M}_\odot) = 11.2$. At $z = 0.4$ spiral galaxies were very slightly more clustered with a relative bias 1.05 ± 0.03. Between $0.4 < z < 1.0$ starburst galaxies were very slightly more clustered. Figure 5.19 shows the relative bias between spiral and starburst galaxies with $\log(M_\star/\mathrm{M}_\odot) = 11.2$. The relative bias remained constant over $0.4 < z < 1.0$.

5.3.1.3 Evolution with Redshift

To analyse how the comoving spatial clustering strength of each spectral type has evolved over time we fitted a power law of the form of Equation 2.8 ($r_0(z) = r_0\ (1+z)^{1-(3+\epsilon)/\gamma}$,

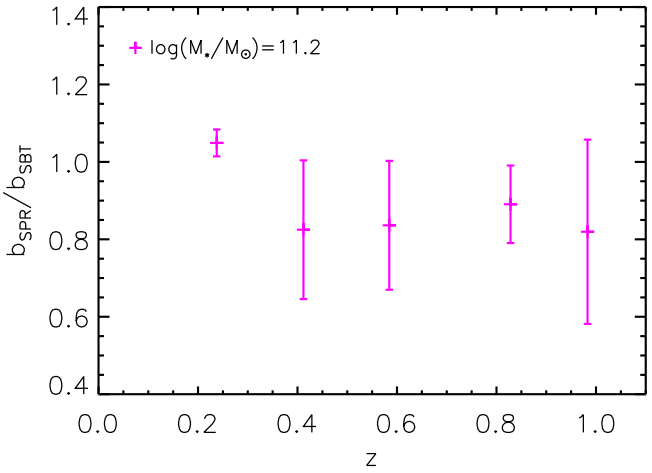


Figure 5.19: The relative bias, calculated using Equation 5.8, between spiral and starburst galaxies with $\log(M_\star/\mathrm{M}_\odot) = 11.2$ over $0.2 < z < 1.0$.

Redshift	$(b_{ELP}/b_{SPR})_{11.6}$	$(b_{ELP}/b_{SPR})_{11.2}$	$(b_{SPR}/b_{SBT})_{11.2}$
0.2	-	1.18 ± 0.13	-
0.4	-	1.26 ± 0.04	1.05 ± 0.03
0.6	-	1.45 ± 0.18	0.82 ± 0.18
0.8	1.44 ± 0.12	-	0.84 ± 0.17
1.0	1.52 ± 0.25	-	0.89 ± 0.10
1.2	-	-	0.82 ± 0.24

Table 5.2: The relative bias (Equation 5.8) of the comoving spatial clustering strength between elliptical (ELP) and spiral (SPR) galaxies with stellar masses $\log(M_\star/\mathrm{M}_\odot) = 11.6$ and $\log(M_\star/\mathrm{M}_\odot) = 11.2$. Also the relative bias between spiral and starburst galaxies (SBT) with stellar masses $\log(M_\star/\mathrm{M}_\odot) = 11.2$ were calculated.

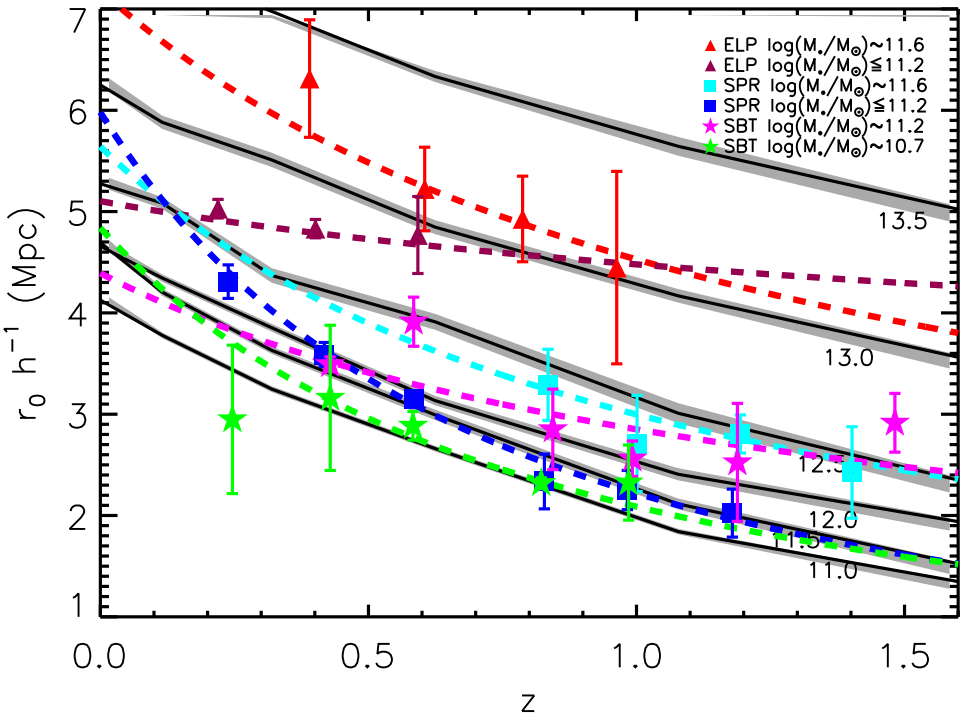
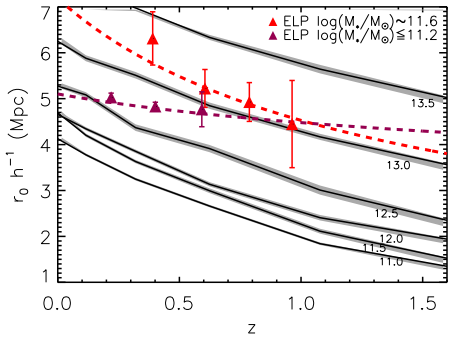
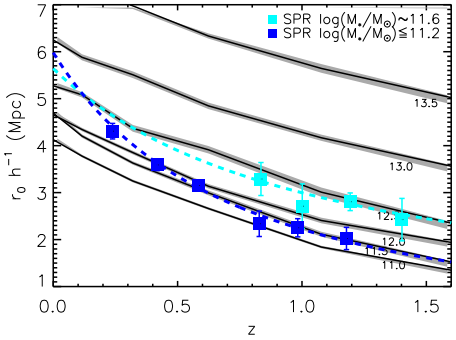


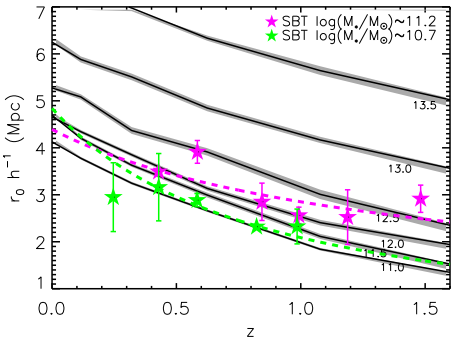
Figure 5.20: The comoving clustering strength r_0 as a function of redshift for elliptical (ELP), spiral (SPR) and starburst (SBT) galaxies plotted as triangles, squares and star symbols, respectively. For ellipticals and spirals, low mass cells ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) were averaged together as such cells showed no variation in clustering strength with stellar mass. Different stellar masses are indicated by the legend. Dashed lines are power law fits (Equation 2.8). Solid black lines with grey error bars show the evolution of the clustering strength for dark matter haloes with $\log(M_{h0}/\mathrm{M}_\odot)$ =11.0, 11.5, 12.0, 12.5, 13.0, 13.5 from the Millennium Simulation. Figure 5.21 shows this plot broken down by type and stellar mass.



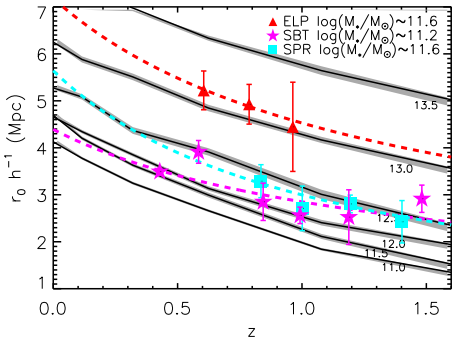
(a) Elliptical galaxies



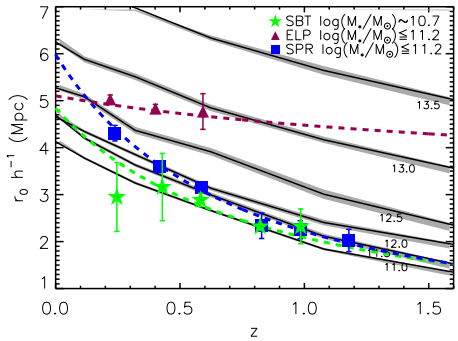
(b) Spiral galaxies



(c) Starburst galaxies



(d) High mass galaxies



(e) Low mass galaxies

Figure 5.21: Comoving clustering strength for elliptical, spiral and starburst galaxies. For ellipticals and spirals, low mass cells ($\log(M_\star/M_\odot) \leq 11.2$) were averaged together as such cells showed no variation in clustering strength with stellar mass. $r_0 - z$ power law fits are plotted as dashed lines. Panels a), b) and c) show just the elliptical, spiral and starburst galaxies, respectively. Panels d) and e) show the high and low stellar mass ranges for each galaxy type. Solid black lines with grey error bars show the evolution of the clustering strength for dark matter haloes with $\log(M_{h0}/\mathrm{M}_\odot)$ =11.0, 11.5, 12.0, 12.5, 13.0, 13.5 from the Millennium Simulation, see Section 5.4.

Galaxy type	$\log(M_\star/\mathrm{M}_\odot)$	ϵ	r_0 (h^{-1}Mpc)
Elliptical	11.6	-0.01 ± 1.55	7.18 ± 3.32
Elliptical	11.2	-0.86 ± 0.89	5.10 ± 0.88
Spiral	11.6	0.44 ± 1.11	5.64 ± 2.60
Spiral	11.2	1.37 ± 0.32	5.98 ± 0.45
Starburst	11.2	-0.08 ± 0.25	4.39 ± 0.27
Starburst	10.7	0.98 ± 0.54	4.84 ± 0.84

Table 5.3: Parameters ϵ and r_0 for a power law fit of the form of Equation 2.8 quantifying the evolution of the comoving spatial clustering strength with redshift for elliptical, spiral and starburst galaxies. Fits were made separately for the stellar mass ranges indicated.

setting γ=1.8) to the clustering strengths. The clustering strength of elliptical and spirals was invariant with stellar mass over $\log(M_\star/\mathrm{M}_\odot) \leq 11.2$, therefore the low mass cells were averaged together at each redshift. We fitted separate power laws for $\log(M_\star/\mathrm{M}_\odot) \leq 11.2$ and $\log(M_\star/\mathrm{M}_\odot) = 11.6$ mass galaxies. For starburst galaxies separate power laws were fitted for the two stellar mass bins $\log(M_\star/\mathrm{M}_\odot) = 10.7$ and $\log(M_\star/\mathrm{M}_\odot) = 11.2$. See Table 5.3 for best-fitting ϵ and r_0 parameters and Figure 5.21 for the plotted power laws. Figure 5.20 shows r_0 as a function of redshift for elliptical, spiral and starburst galaxies along with the power laws. Figures 5.21(a), 5.21(b) and 5.21(c) show the same results but with elliptical, spiral and starburst galaxies plotted separately. Figures 5.21(d) and 5.21(e) show the high and low stellar mass ranges for each type.

For elliptical galaxies with $\log(M_\star/\mathrm{M}_\odot) = 11.6$ the clustering strength decreased with z with $\epsilon = -0.01 \pm 1.55$. However, given the errors on the r_0 values a constant clustering strength could not be ruled out. The clustering strength of low mass ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) elliptical galaxies decreased much more slowly with redshift. The clustering strength of $\log(M_\star/\mathrm{M}_\odot) \leq 11.2$ spiral galaxies showed a strong decrease with redshift with $\epsilon = 1.37 \pm 0.32$. The large redshift range probed by these galaxies and small errors on r_0 meant the change in clustering strength was well constrained. High mass spiral galaxies ($\log(M_\star/\mathrm{M}_\odot) = 11.6$) showed a decrease in r_0 with z over $0.8 < z < 1.4$. The clustering of starburst galaxies was similar to spiral galaxies except at low redshift where they were more weakly clustered leading to the clustering strength of starburst galaxies showing less change with redshift.

5.3.2 SFR-Redshift Results

The clustering of spiral and starburst galaxies (collectively) selected on SFR was measured in five redshift bins over $0.1 < z < 1.5$. Two bins measured how the clustering strength varied with SFR at $z = 0.3$ with the low SFR bin sampling sub-LIRG galaxies and high SFR bin containing LIRG galaxies. The redshift bin at $z = 0.71$ also contained LIRGS and the bins at $z = 1.06$ and 1.38 contained ULIRGS.

The angular correlation function in each cell was modeled as a power law (Equation 2.19) with no evidence for any evolution in γ. Figures 5.22 and 5.23 show the angular correlation functions for ELAIS-N1 and ELAIS-N2, respectively, for each $(SFR - z)$ cell. The clustering amplitudes A along with the comoving spatial clustering strengths r_0 for each cell are shown in Table 5.10.

5.3.2.1 Variation of Clustering Strength with SFR

The two cells at $z = 0.3$ with mean $\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) = 0.5$ and $\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) = 1.3$ had comoving clustering strengths $r_0 = 3.18 \pm 0.31\ h^{-1}$Mpc and $r_0 = 3.16 \pm 0.23\ h^{-1}$Mpc, respectively. Therefore the clustering strength was invariant with SFR at this redshift over the range of star formation rates probed. This implies these star forming galaxies are found in the same environments regardless of their SFR. Table 5.10 shows stellar mass estimates for the (SFR, z) cells. The low and high SFR cells had stellar masses of $\log(M_{\star}/\mathrm{M}_{\odot}) = 10.94$ and 11.14, respectively. These stellar masses are both in the low mass regime for which no variation of clustering strength was seen in the $(M_{\star}, z)$ analysis and hence might also explain why no variation with SFR was seen.

5.3.2.2 Evolution with Redshift

The comoving spatial clustering strength increased from $r_0 = 3.16 \pm 0.23\ h^{-1}$Mpc at $z = 0.3$ to $4.24 \pm 0.85\ h^{-1}$Mpc at $z = 1.38$, see Figure 5.24. The mean SFR of each bin increased with redshift therefore it was difficult to draw any definitive conclusions about the evolution of the clustering. Ideally the clustering would be measured for galaxies of the same SFR across a range of redshifts. The two cells at $z = 0.3$ showed the clustering strength to be invariant with SFR but this can only be assumed to be true at this redshift for the SFR range probed.

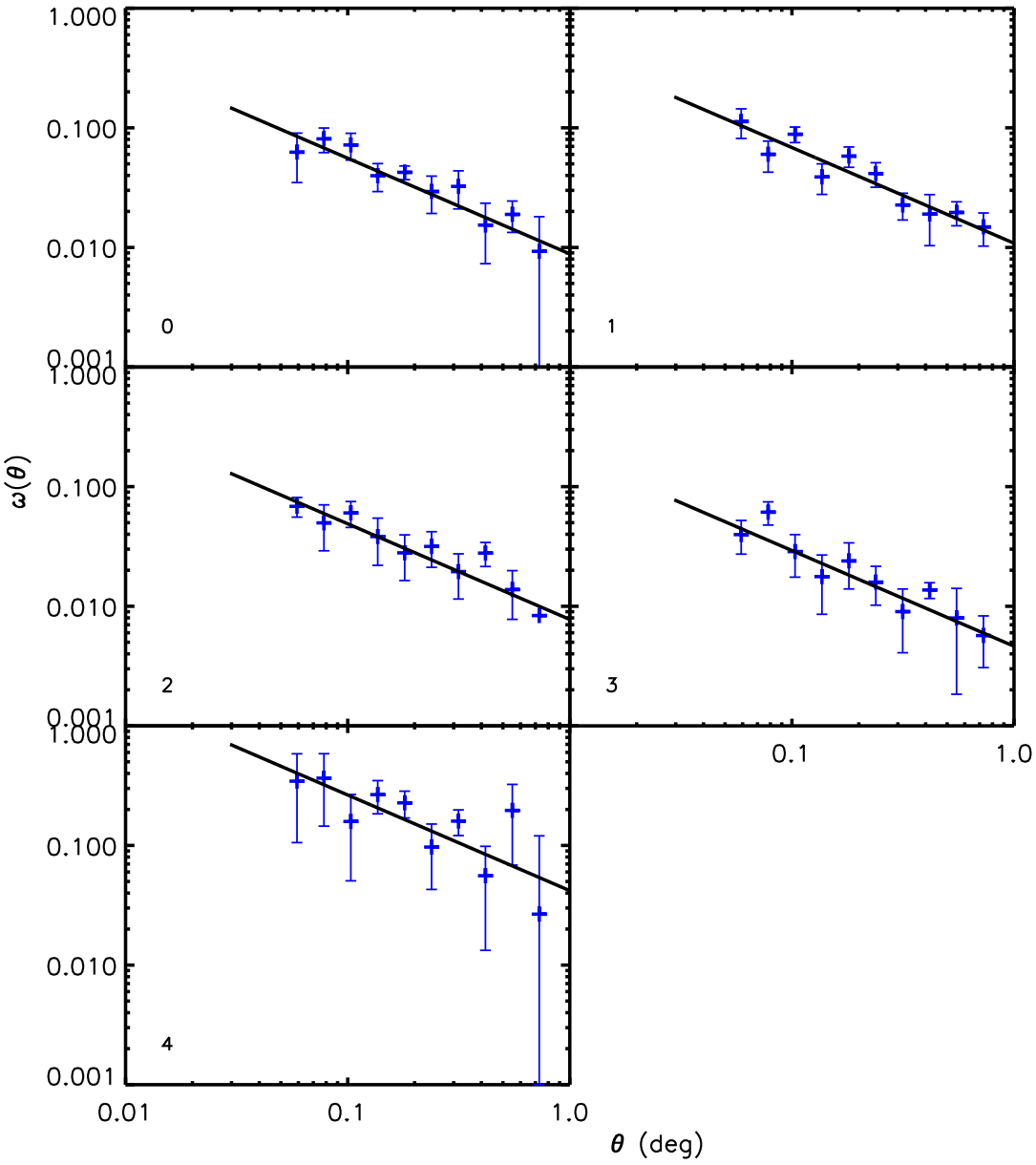


Figure 5.22: The angular correlation function, $\omega(\theta)$, for SFR selected galaxies in ELAIS-N1. The power law fit was made with $\gamma = 1.8$. Each panel is for a different (SFR, z) cell, indicated by the number in the panel, see Table 5.10 for SFR and redshift ranges.

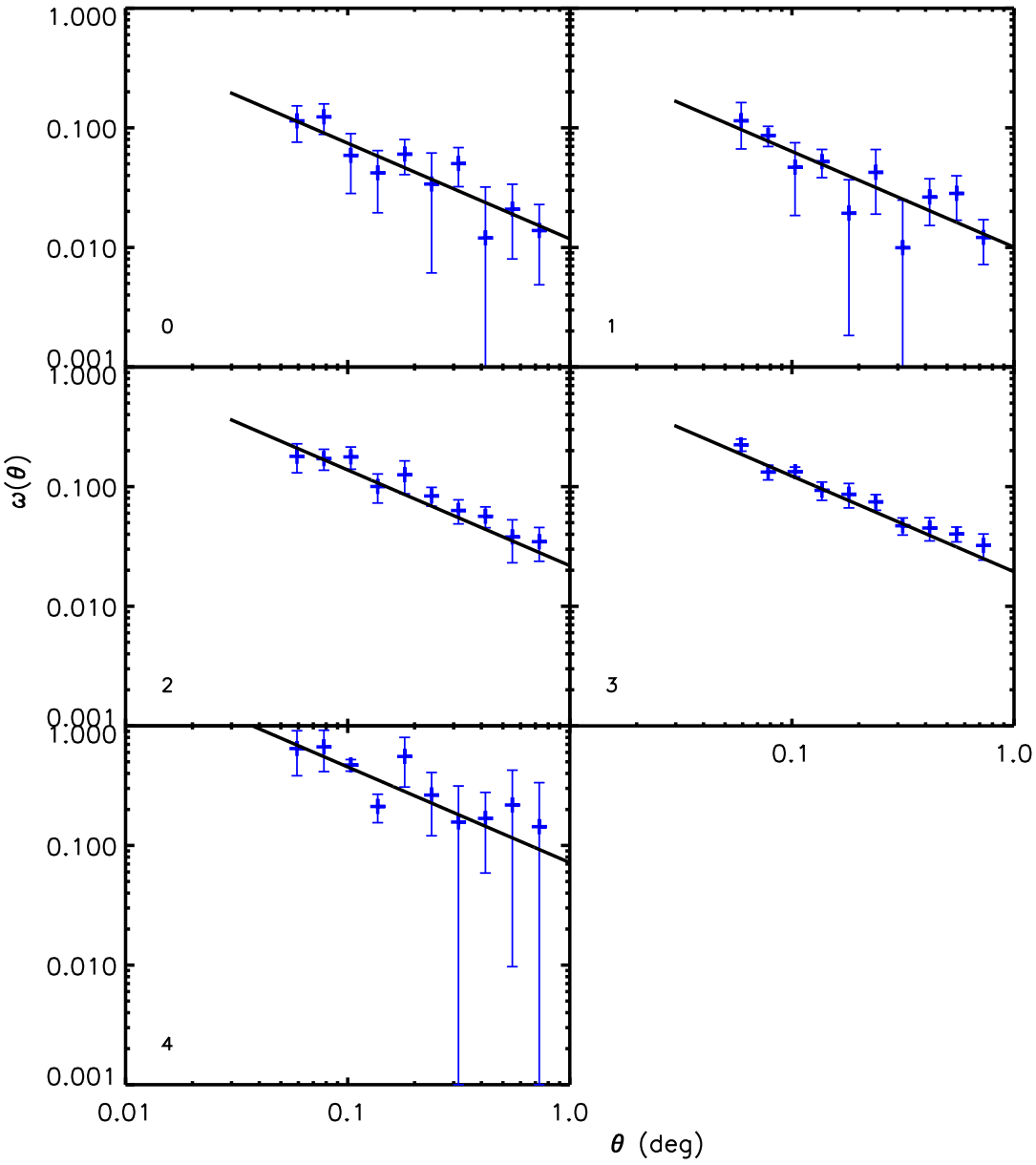


Figure 5.23: The angular correlation function, $\omega(\theta)$, for SFR selected galaxies in ELAIS-N2. The power law fit was made with $\gamma = 1.8$. Each panel is for a different (SFR, z) cell, indicated by the number in the panel, see Table 5.10 for SFR and redshift ranges.

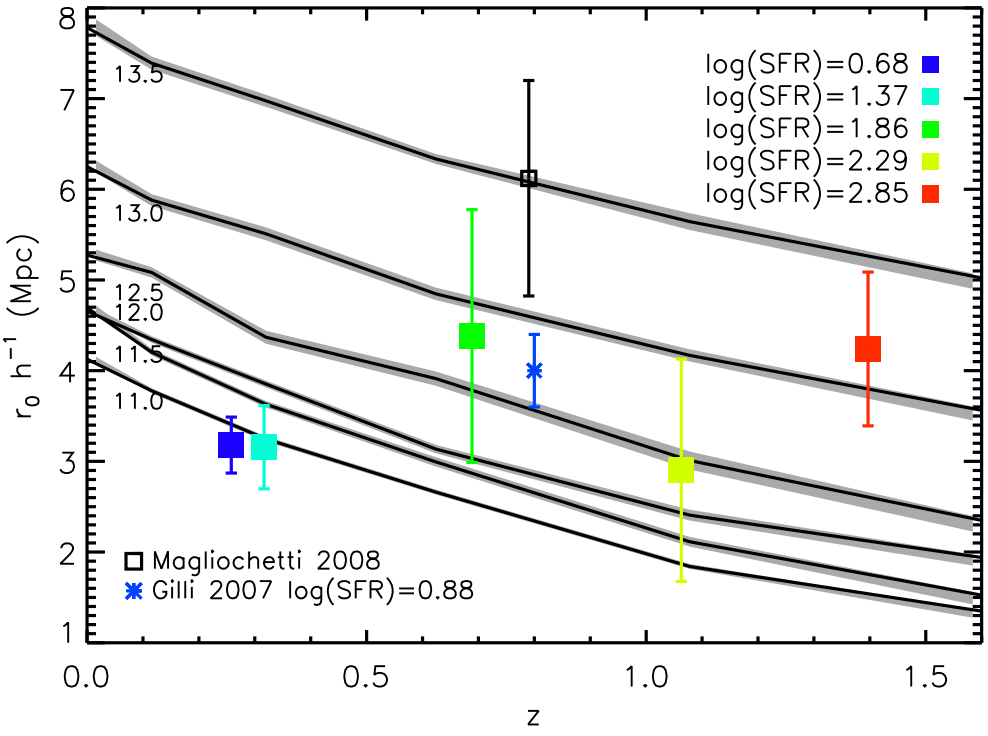


Figure 5.24: The comoving clustering strength r_0 as a function of redshift and SFR. Square data points are for this analysis. Additional data points from the literature are Magliocchetti et al. (2008) and Gilli et al. (2007). Each data point is coloured to represent the SFR indicated by the legends in units of $M_{\odot}yr^{-1}$(Magliochetti et. al. (2008) is shown in black as it has no SFR estimate).

5.4 Comparison to Simulated Dark Matter Haloes

Measuring the clustering of galaxies provides a way to indirectly measure the clustering of the underlying dark matter distribution. Galaxies are said to be biased tracers of the dark matter distribution with different types and/or luminosities thought to have different biases. In order to further our understanding of galaxy evolution and the typical environments galaxies reside in, it is important to determine typical parent halo masses for the SWIRE galaxies.

A common method used to determine parent halo masses is the halo occupation distribution (HOD) model (e.g. Kauffmann et al.(1997), Peacock & Smith(2000)). The HOD model describes the number and distribution of galaxies within dark matter haloes and hence concentrates on small scale power. Other analytic approaches include matching the galaxy luminosity function with the predicted halo mass function (e.g. Yang et al.(2003) and Moster et al.(2009)). Here we follow the approach of Gilli et al.(2007) and Giavalisco & Dickinson(2001) to determine parent halo masses by comparing the clustering strengths (r_0) of SWIRE galaxies with the clustering strengths of simulated dark matter haloes. We use our measurements of the clustering of haloes in the Millennium Simulation from Chapter 4. The clustering of SWIRE galaxies and Millennium Simulation haloes were both measured at large inter-halo separations in comoving coordinates. At large separations the galaxy clustering is thought to follow the dark matter thereby giving this method validity. The clustering of dark matter haloes from the Millennium Simulation were measured for two different analyses:

- *Fixed mass clustering analysis:* The clustering of haloes of masses $\log(M_\star/\mathrm{M}_\odot)$ =10.5, 11.0, 11.5, 12.0, 12.5, 13.0 and 13.5 were measured in comoving coordinates at select redshifts over $0 < z < 1.5$, i.e. the same mass haloes were selected at each redshift (for the method used see Section 4.4.4).

- *Evolution of clustering analysis:* Haloes of masses $\log(M_\star/\mathrm{M}_\odot)$ =11.0, 11.5, 12.0, 12.5, 13.0 and 13.5 at $z = 0$ were selected and the progenitors of these haloes on the main branch of their merger were also selected at certain redshifts out to $z = 1.5$. The clustering of the $z = 0$ haloes of each mass was measured as was the clustering of the progenitors in each mass range at higher redshifts, i.e. the same haloes were selected at each redshift enabling the evolution of the clustering of the dark matter to be measured.

The SWIRE galaxy r_0 values were compared to the r_0 values for the dark matter haloes from the *fixed mass* and *evolution* analyses. Comparing to the fixed mass analysis enabled the typical mass of the parent haloes, denoted M_{hz}, which the galaxies resided in at their current redshift to be determined for each cell in the $M_\star - z$ and $SFR - z$ analyses, see Figures 5.25(a) and 5.26(a), respectively. Comparing to the evolution analysis enabled the typical mass halo that the galaxies will reside in at redshift zero, denoted M_{h0}, to be determined for both analyses, see Figures 5.25(b) and 5.26(b). Hence this will indicate how the clustering strength of the galaxies will change over time and inferences can be made on the sort of environments the galaxies will reside in at later times.

The typical M_{hz} and M_{h0} were determined for each $(M_\star, z)$ and (SFR, z) cell by interpolating over halo mass to determine the halo mass which would have the same clustering strength as the galaxies in the cell. Note, halo masses were determined for the low stellar mass ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) cells averaged together and not for each individual cell, although they would be approximately the same. Tables 5.5, 5.7 and 5.9 show the typical M_{hz} and M_{h0} for the averaged low mass $(M_\star, z)$ cells and high mass $(M_\star, z)$ cells for elliptical, spiral and starburst galaxies, respectively. Table 5.10 shows M_{hz} and M_{h0} for the (SFR, z) cells.

5.4.1 Parent Halo Masses

Figure 5.25(a) shows the typical halo mass for elliptical and spiral galaxies as a function of redshift in each $(M_\star, z)$ cell. At $z = 1.0$ elliptical galaxies of mass $\log(M_\star/\mathrm{M}_\odot) = 11.6$ were typically found in $\log(M_{hz}/\mathrm{M}_\odot) = 12.2^{+0.4}_{-0.9}$ haloes increasing to $12.6^{+0.1}_{-0.3}$ at $z = 0.6$ meaning that such galaxies have become more biased with respect to the dark matter. The less massive ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) elliptical galaxies at $z = 0.2$, $z = 0.4$ and $z = 0.6$ were typically found in $\log(M_{hz}/\mathrm{M}_\odot) \sim 12.3$ haloes indicating they were found in similar environments at their respective redshifts. Therefore the bias of low mass elliptical galaxies has not changed much over this redshift range. The typical parent halo of massive ($\log(M_\star/\mathrm{M}_\odot) = 11.6$) spiral galaxies increased from $\log(M_{hz}/\mathrm{M}_\odot) = 10.4^{+0.4}_{-0.4}$ to $10.9^{+0.6}_{-0.4}$ over $0.8 < z < 1.4$. The typical parent halo mass of low mass ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) spiral galaxies increased from $\log(M_{hz}/\mathrm{M}_\odot) \sim 10.0$ at $z = 0.8$ to 11.7 at $z = 0.2$ indicating such galaxies are much more biased tracers of the dark matter at low redshifts. Starburst galaxies of mass $\log(M_\star/\mathrm{M}_\odot) = 11.2$ were typically found in $\log(M_{hz}/\mathrm{M}_\odot) \sim 10.7$ haloes with a large amount of scatter. Lower mass starburst galaxies ($\log(M_\star/\mathrm{M}_\odot) = 10.7$) were

found in $\log(M_{hz}/\mathrm{M}_{\odot}) \sim 10.2$ haloes.

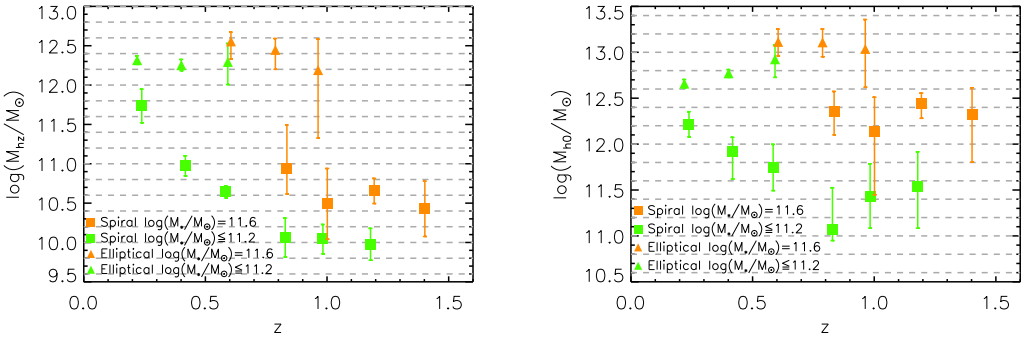
Figure 5.26(a) shows the typical parent halo mass for the galaxies in each (SFR, z) cell. The galaxies at $z = 0.3$ were housed in haloes of $\log(M_{hz}/\mathrm{M}_{\odot}) \sim 11.0$ compared to 13.1 for the high redshift star forming galaxies at $z = 1.4$. This is evidence for downsizing in halo mass with star formation progressively found in lower mass haloes as we move toward $z = 0$.

The parent halo masses inferred for spiral and starburst galaxies are too low, particularly at high redshift. Situations where galaxies were inferred to have halo masses smaller than the stellar mass of the galaxy is physically implausible. However, spiral and starburst galaxies which were found in $\log(M_{hz}/\mathrm{M}_{\odot}) \sim 10.0$ haloes were very close to the mass resolution limit of the Millennium Simulation which is about $2.0 \times 10^{10} M_{\odot}$. Therefore the inferred halo masses should be treated with caution. Perhaps more importantly at low halo masses the rate of change of clustering strength with halo mass is very small, see Figure 4.5. Therefore it would only have required a small increase in clustering strength to infer a much larger halo mass. Also our method assumed that the clustering of haloes was dependent solely on their mass, but more clustered haloes does not necessarily mean they are more massive. Work by Gao et al.(2005), Zhu et al.(2006), Wechsler et al.(2006) and Croton et al.(2007) have shown that the clustering of dark matter haloes is also dependent on formation redshift, (as discussed in Section 4.4.5). Our method to assign halo masses did not take this into account. We conclude that this method is good for determining *typical* parent halo masses, particularly for higher mass galaxies with larger biases, but that it doesn't work well for lower mass galaxies.

5.4.2 Evolution of Dark Matter Halo Clustering

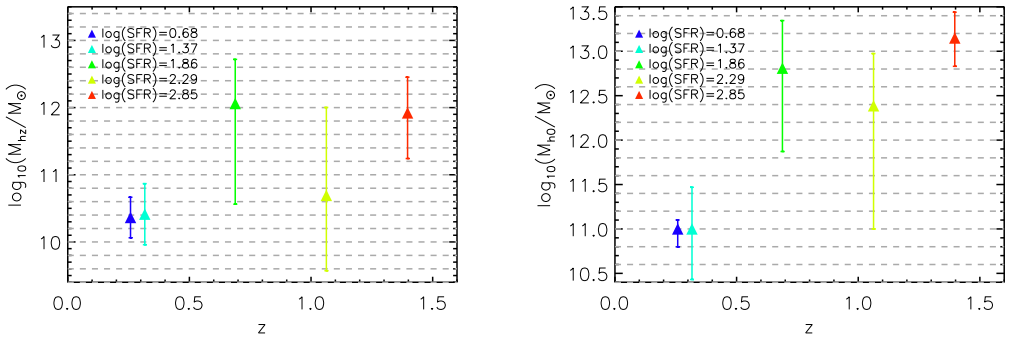
We used the evolution of the comoving clustering strength of haloes over $0 < z < 1.5$ as *evolutionary tracks* to indicate the sort of environments the galaxies will reside in at later times. Section 5.4.1 brought into question the accuracy of the parent halo masses, particularly for low mass haloes. However the evolution of the clustering strength of the dark matter haloes evolves at the same rate regardless of halo mass. Therefore the evolutionary tracks can still be used to infer how the clustering of galaxies will evolve without worrying about the exact mass of the haloes.

Figure 5.25(b) shows the M_{h0} as a function of redshift. The high mass ($\log(M_{\star}/\mathrm{M}_{\odot}) = 11.6$) elliptical galaxies traced the evolution of the clustering of $\log(M_{h0}/\mathrm{M}_{\odot}) = 13.1$



(a) Typical parent halo mass, M_{hz} (b) Typical halo mass at $z = 0$, M_{h0}

Figure 5.25: Halo masses for $\log(M_\star/\mathrm{M}_\odot) \leq 11.2$ and $\log(M_\star/\mathrm{M}_\odot) = 11.6$ elliptical and spiral galaxies in each $(M_\star, z)$ cell. Plot a) shows the typical halo mass the galaxies reside in at their current redshift, M_{hz}. Plot b) shows the typical halo mass they will reside in at redshift zero, M_{h0}. Galaxies along the same horizontal lines would be found in the same sort of environments at the present day. Data points are colour coded by $M_\star$ indicated in the legend.



(a) Typical parent halo mass, M_{hz} (b) Typical halo mass at $z = 0$, M_{h0}

Figure 5.26: Halo masses for spiral galaxies in each (SFR, z) cell. Plot a) is shows the typical halo mass the galaxies reside in at their current redshift, M_{hz}. Plot b) shows the typical halo mass they will reside in at redshift zero, M_{h0}. Galaxies along the same horizontal lines would be found in the same sort of environments at the present day. Data points are colour coded by SFR as indicated by the legend in units of $\mathrm{M}_\odot\mathrm{yr}^{-1}$.

haloes closely. These galaxies at $z = 1.0$ are in the sort of haloes which will evolve by $z = 0.6$ to be in the sort of environments which house similar mass galaxies at later times. Such galaxies will reside in groups or small clusters at $z = 0$. The low and high mass spiral galaxies at $z \geq 0.8$ traced the evolution of the halo clustering quite well. The low mass spiral galaxies detected at $0.2 < z < 0.6$ were progressively more biased with respect to the dark matter approaching the present day. The low mass elliptical galaxies detected at $z = 0.6, 0.4, 0.2$ will reside in $\log(M_{h0}/\mathrm{M}_{\odot}) = 12.9, 12.8, 12.7$ at $z = 0$, respectively.

Comparing Figures 5.25(b) and 5.26(b) we can see the highly star forming galaxies ($\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) = 2.85$) at $z = 1.4$ are in the sort of environments which will have evolved by $z \sim 0.6 - 0.8$ to house $\log(M_{\star}/\mathrm{M}_{\odot}) \sim 11.6$ elliptical galaxies. This implies that these high star forming galaxies will have their star formation quenched by $z \sim 0.6 - 0.8$ and will evolve into ellipticals.

5.5 Discussion

The angular correlation function was reliably measured for all cells for both analyses. This reflects the quality of the criteria which we defined to ensure the selections effects were satisfactorily constrained. We found no change in γ (the gradient of the correlation function) between ellipticals and spirals nor any variation with stellar mass. We attribute this to the fact that we measured the clustering at large scales ($\gtrapprox 3$ Mpc) where the inter-halo clustering was probed. Therefore our analyses were not sensitive to the small-scale intra-halo clustering. On such scales γ probes the radial profile of the dark matter haloes and has been shown to be larger for ellipticals and vary with luminosity in some analyses, e.g. Zehavi et al.(2004), Madgwick et al.(2003).

There are few clustering strengths quoted in the literature for galaxy samples constrained by stellar mass or SFR, and spectral type at high redshift. The difference in clustering between elliptical and spiral galaxies has been shown to be largely insensitive to how the two populations are separated, whether by spectral or morphological type or rest-frame colour (e.g. Norberg et al.(2002), Madgwick et al.(2003), Coil et al.(2004)). Therefore we compared the results from the $M_{\star} - z$ analysis to comoving clustering strengths from the literature determined for elliptical/early-type/red and spiral/late-type/blue galaxies. Very few clustering analyses have looked at the variation with stellar mass and instead have tended to concentrate on luminosity. Therefore Tables 5.4, 5.6 and 5.8 quote the

B-band luminosity, L_B, for each cell as an indicator of the luminosity of the cells.

We found elliptical galaxies were more clustered than spiral galaxies at all redshifts. This shows that elliptical galaxies were preferentially found in denser environments out to $z = 1.0$ indicating the colour-density relation existed at these earlier times in agreement with Coil et al.(2008) and McCracken et al.(2008). The inferred dark matter halo masses support this conclusion by showing the typical parent halo mass for ellipticals ($\log(M_{hz}/\mathrm{M}_\odot) \sim 12.5$) to be much greater than that of spirals ($\log(M_{hz}/\mathrm{M}_\odot) \sim 11.0$), although these should be treated with caution.

The relative bias between ellipticals and spirals was in broad agreement with estimates from the literature which are typically in the range 1.0-1.7 (Willmer et al.(1998), Madgwick et al.(2002), Madgwick et al.(2003), Coil et al.(2008)), although the exact calculation and scales used to measure the relative bias varies from one study to another. Shepherd et al. (2001) found a relative bias of 1.7 ± 0.2 at $z = 0.4$ which does not agree with our value of 1.26 ± 0.05 at this redshift. However, Shepherd et al. (2001) measured the bias at smaller scales, where the relative bias has been shown to increase (e.g. Coil et al.(2008), Madgwick et al.(2003)). Also Shepherd et al. (2001)'s sample was not constrained by stellar mass. If our sample incorporated galaxies of all mass, our spiral sample would not change as we did not detect more massive spirals at this redshift but incorporating the high mass ellipticals with their stronger clustering strength would cause the relative bias to increase. At $z \sim 0.8$ Coil et al.(2008) found the relative bias between red and blue galaxies to be 1.28 ± 0.09 (at large scales). This is in good agreement with our value of 1.44 ± 0.12 at the same redshift.

We could not conclude whether the relative bias between elliptical and spiral galaxies increased with redshift or remained constant. The relative bias of $\log(M_\star/\mathrm{M}_\odot) = 11.2$ galaxies at $0.2 \leq z \leq 0.6$ showed an increase with increasing redshift but could be considered to be constant to within errors. $\log(M_\star/\mathrm{M}_\odot) = 11.6$ galaxies showed no change over $0.8 \leq z \leq 1.0$. By comparison Coil et al.(2008) found a constant relative bias over $0 < z < 1.0$. If the relative bias has not changed over $0 < z < 1.0$ then this shows the colour-density relation was as strong at $z = 1$ as it is today. Coil et al.(2008) state that linear growth theory predicts that the relative bias of our elliptical and spiral samples should decrease from $z = 1$ to $z = 0$ as less dense regions show more growth at late times. It is well known that a fraction of spiral galaxies have turned into elliptical galaxies between $z = 1$ and $z = 0$, Lilly et al.(1995), Cowie et al.(1996), Gabasch et al.(2004), Bell et

al.(2004). Therefore if the relative bias is expected to decrease between $z = 1$ and $z = 0$ this would require the most clustered spiral galaxies to be turning into ellipticals in order to produce a constant relative bias.

The clustering of elliptical and spiral galaxies was invariant with stellar mass at low masses ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$), in agreement with other studies performed over a similar redshift range e.g. Coil et al.(2008), McCracken et al.(2008) although perhaps at odds with Meneux et al.(2009) who found a very weak mass dependence at $0.2 < z < 1.0$. However Meneux et al.(2009) did not split their sample into early and late-type galaxies, hence the increasing relative mix of early-to-late-types with stellar mass would amplify the clustering signal. We attribute the clustering invariance with stellar mass to low stellar mass galaxies being found in haloes with a larger range of masses as they are less biased with respect to the dark matter.

For the first time, we measured the clustering strength at $z \geq 0.4$ of high stellar mass galaxies ($\log(M_\star/\mathrm{M}_\odot) > 11.2$) where we found a strong dependence on stellar mass content. This is in agreement with low redshift measurements who probed to $L > L^*$ luminosities e.g. Norberg et al.(2002), Budavári et al.(2003) and Shepherd et al. (2001). High stellar mass galaxies are more strongly biased tracers of the dark matter distribution. This is reflected in the inferred halo masses which were larger for the more massive galaxies of each type. Therefore such galaxies showed a stronger in clustering strength.

Spiral galaxies of all stellar masses showed a decrease in clustering strength with increasing redshift. At $z > 0.8$ spiral galaxies traced the evolution of the halo clustering quite well. We attribute this to these galaxies continuously forming stars at a steady rate and hence their luminosity and hence mass didn't change much over time. At $z < 0.8$ low mass spirals ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) showed a strong increase in clustering strength with decreasing redshift. Therefore at high redshift they were found in less dense environments and over time such galaxies became progressively more biased tracers of the dark matter eventually being found in denser environments at low redshift. Elliptical galaxies of mass $\log(M_\star/\mathrm{M}_\odot) \leq 11.2$ showed very little change in clustering strength with redshift. High mass ellipticals ($\log(M_\star/\mathrm{M}_\odot) = 11.6$) showed a larger decrease with z but a constant clustering strength could not be ruled out due to large errors on r_0. Starburst galaxies showed very little change in clustering strength with redshift. We conclude the clustering of elliptical galaxies has not changed much over $0 < z < 1$ whereas low mass spiral galaxies have shown a large change. We conclude the densest regions of the Universe are already

in place by $z \sim 1$ and the large scale structure in less dense regions continue to grow at later times.

We compare how our comoving spatial clustering strengths vary with redshift to values from the literature. Waddington et al.(2007) measured the clustering of SWIRE galaxies selected on 3.6μm flux. They fitted the same form of equation (Equation 2.8) to their comoving clustering strengths and found a much stronger decrease with redshift with $\epsilon = 4.8 \pm 0.8$. However, they were fitting to galaxies of all types and stellar masses at $0.45 < z < 1.0$. Over this redshift range their sample showed a larger variation partially due to the change in the relative mix of ellipticals-to-spirals which decreased with redshift. Unfortunately this is indicative of many clustering analyses which have attempted to quantify the evolution with redshift. As outlined in Section 1.3.3, previous clustering analyses such as de la Torre et al.(2007), Le Fèvre et al.(2004), Meneux et al.(2008), Meneux et al.(2009) have been constrained to galaxy samples selected based on observable properties leading to varying intrinsic properties with redshift or mixing spectral types making it impossible to separate the contribution to the clustering signal from early and late-type galaxies.

McCracken et al.(2008) investigated the evolution of the clustering of early and late-type galaxies over $0.4 < z < 1.2$ selected on B band absolute magnitude, M_B. Their early-type galaxies had median $M_B \sim -21.0$. They used photometric redshifts to bin their galaxies into 5 redshift bins over $0.4 < z < 1.4$. Their r_0 values as a function of z are shown in Figure 5.27. The clustering of their early-type galaxies showed a small decrease with redshift. Our high mass ($\log(M_\star/\mathrm{M}_\odot) = 11.6$) elliptical galaxies at $z = 0.6, 0.8, 1.0$ also showed a small decrease with z. Our elliptical galaxies were consistently more clustered but did agree to within errors at at $z = 0.8$ and $z = 1.0$. This was expected as our high mass elliptical galaxies typically had $M_B \sim 22.0$ and hence were more luminous than those of McCracken et al.(2008).

The clustering strength of McCracken et al.(2008)'s late-type galaxies, with $M_B = -20.1$, showed no variation with redshift whereas our low mass ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) spiral galaxies showed a large change in r_0 over $0.4 < z < 0.8$ but little change over $0.8 < z < 1.2$. At $z \geq 0.6$ there was good agreement between the clustering of our low mass spirals and McCracken et al.(2008)'s late-type galaxies despite our galaxies being more luminous with $M_B \sim -21.4$. This reflects the trend, seen in both our analyses, that at low luminosities (or stellar mass) r_0 did not change with luminosity. At $z = 0.4$

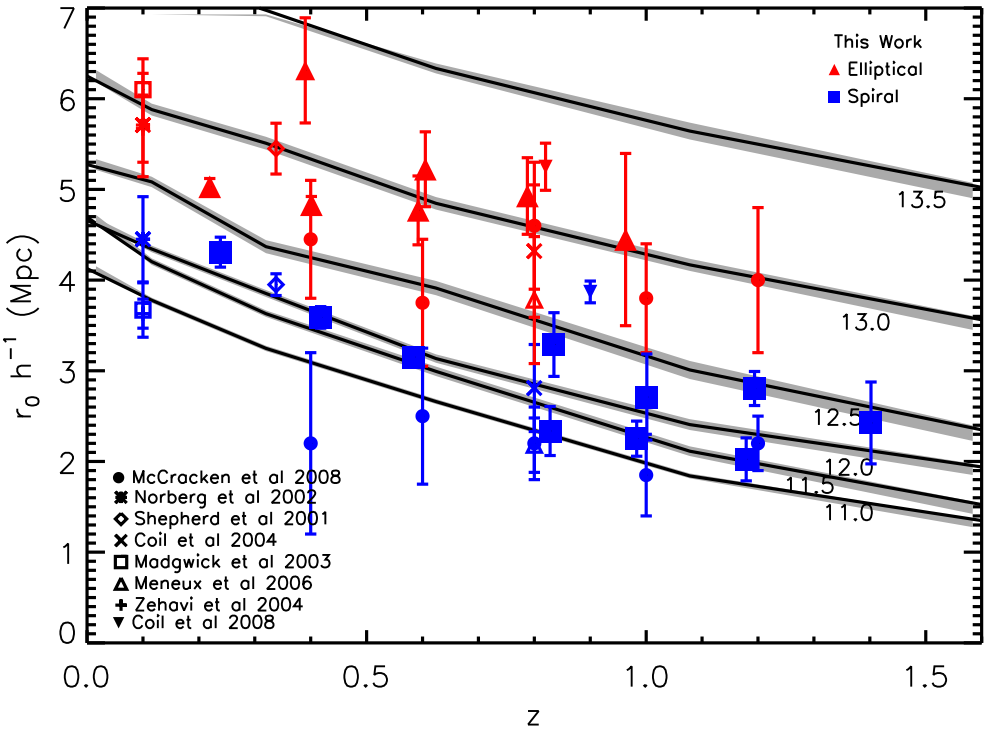


Figure 5.27: The comoving clustering strength r_0 as a function of redshift for elliptical/early-type/red galaxies (red points), spiral/late-type/blue galaxies (blue points). Starburst galaxies are removed for clarity. Elliptical and spiral galaxies from this analysis are plotted as triangles and squares, respectively. r_0 values from Norberg et. al. 2002, Shepherd et. al. 2001, Madgwick et. al. 2003, Coil et. al. 2004, Zehavi et. al. 2004, Meneux et. al. 2006, McCracken et. al. 2008 and Coil et. al. 2008 are also plotted, symbols are indicated in the legend.

our spiral galaxies were more strongly clustered however the clustering strength of our $\log(M_\star/\mathrm{M}_\odot) = 10.7$ starburst galaxies was in agreement with McCracken et al.(2008)'s late-type galaxies.

Figure 5.27 shows the comoving spatial clustering strengths from the $(M_\star, z)$ analysis along with those determined by Norberg et al.(2002), Shepherd et al. (2001), Madgwick et al.(2003), Coil et al.(2004), Coil et al.(2008), Zehavi et al.(2004), McCracken et al.(2008) and Meneux et al. 2006. It should be noted when comparing values from different surveys care should be taken to consider the physical scale over which the correlation function was measured and the value of γ used in fitting a power law to the correlation function. Due

to such differences between surveys detailed comparisons are difficult to perform. The clustering strength of our elliptical galaxies at $z = 0.2$ and $z = 0.4$ are smaller than those determined by Norberg et al.(2002), Madgwick et al.(2003) and Shepherd et al. (2001) at similar redshifts. However these literature values were determined for $L > L^*$ galaxies which is beyond the luminosity range of our analysis at these redshifts. At $z = 0.8$ our most massive elliptical galaxies ($L > L^*$) have a similar clustering strength to Coil et al.(2008) optically selected red galaxies, however their galaxies have $L < L^*$. This difference is probably due to the different scales probed where Coil et al.(2008) probe to smaller scales than our analysis. It might also be a reflection of the different galaxy populations probed by optical and IR surveys. Overall there is good qualitative agreement between our results and other similar analyses from the literature.

The clustering of SWIRE galaxies selected on SFR provided evidence for downsizing. At $z = 0.3$ galaxies selected on SFR were much more weakly clustered than those at $z = 1.4$. This was also reflected in a decrease in typical parent halo mass by ~ 2 orders of magnitude. These findings support the assertion of downsizing where star formation is more frequently found in less massive galaxies at low redshift and more massive galaxies at high redshift. However the SFR of the high redshift sample was ~ 100 times larger and hence this might have been partly responsible for the increase in clustering strength. At $z = 0.3$ we found the clustering to be invariant with SFR but this was at $\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1}) < 1.4$.

Here we attempt to determine whether clustering strength increases with SFR at larger SFRs using values from the literature. The clustering strengths of our SFR selected galaxies at $z = 0.71$ are compared to Magliocchetti et al.(2008) and Gilli et al.(2007). Gilli et al.(2007) found at $z \sim 0.8$ star forming galaxies detected at $S_{24\mu\mathrm{m}} > 20\mu$Jy to have $r_0 = 4.0 \pm 0.4\ h^{-1}$Mpc with $\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1}) = 0.9$. Magliocchetti et al.(2008) looked at the clustering of galaxies detected with Spitzer with $S_{24\mu\mathrm{m}} > 400\mu$Jy at $z \sim 0.8$ and found $r_0 = 6.12^{+1.1}_{-1.3}\ h^{-1}$Mpc. Unfortunately no SFR estimate was published with this work, however we estimate a SFR based on that of Gilli et al.(2007). The flux cut used in Magliocchetti et al.(2008) was 20 times larger than that used in Gilli et al.(2007). Therefore we estimate the SFR to be 20 times larger giving $\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1}) = 2.18$. Our (SFR, z) cell at $z = 0.71$ with $\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1}) = 1.86$ had a clustering strength of $r_0 = 4.38 \pm 1.39\ h^{-1}$Mpc. Note, we ignore the small difference in redshift between our value and those from the literature. Figure 5.28 shows these 3 clustering strengths as a

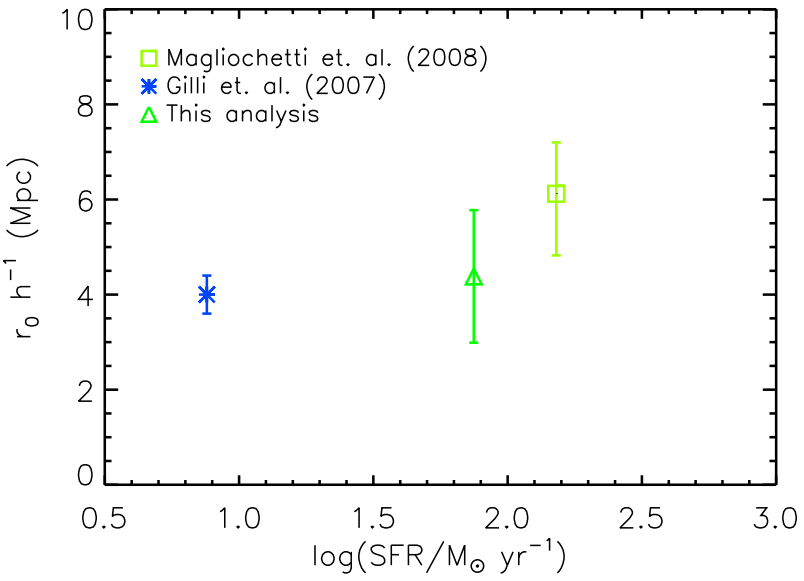


Figure 5.28: The comoving clustering strength as a function of SFR at $z = 0.7$. Additional data points from the literature are Magliocchetti et al. (2008) and Gilli et al. (2007). Each data point is coloured to represent the SFR.

function of SFR. There is no variation in r_0 with SFR at $\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1}) \lesssim 2.0$ but there does appear to be an increase at $\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1}) \gtrsim 2.0$. This implies galaxies with low SFRs are typically found in lower mass haloes and vice versa for galaxies with large SFRs. There are very few clustering analyses in the literature where galaxies have been selected based on SFR. This is an area which requires further investigation.

Comparing our $SFR - z$ analysis with the $M_\star - z$ analysis indicates environment is a more important factor in making galaxies star forming than stellar mass. The (SFR, z) cell at $z = 1.4$ was $\sim 50\%$ more clustered than the $(M_\star, z)$ spiral galaxies at the same redshift. Both samples of galaxies had masses of $\log(M_\star/\mathrm{M}_\odot) = 11.6$. This is also found to be the case at $z = 0.7$ where the SFR selected galaxies were more strongly clustered than spiral galaxies of the same mass. This implies at high redshift that galaxies are star forming because of their environment not because of their stellar mass, i.e. environmental effects are necessary to produce gas instabilities and trigger star formation as opposed to the gas mass being triggered at a given point in the galaxies life almost independently of the environment. At $z = 0.3$ galaxies selected on SFR had a stellar mass of $\log(M_\star/\mathrm{M}_\odot) = 11.0$. Their clustering strength was similar to starburst galaxies with $\log(M_\star/\mathrm{M}_\odot) =$

10.7 and weaker than spiral galaxies of $\log(M_\star/\mathrm{M}_\odot) = 11.2$. This shows at low redshift environment is not an important factor in determining which galaxies are star forming. This is another reflection of downsizing because the bulk of all star formation is taking place in such environments at this time.

The relative bias between spiral and starburst galaxies of mass $\log(M_\star/\mathrm{M}_\odot) = 11.2$ also showed that the starburst galaxies were slightly more clustered at $0.4 < z < 1.0$ indicating the most star forming late-type galaxies are found in denser environments at earlier times.

We found highly star forming galaxies ($\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1}) = 2.85$) at $z = 1.4$ are in the sort of environments which will have evolved by $z \sim 0.6-0.8$ to house $\log(M_\star/\mathrm{M}_\odot) \sim 11.6$ elliptical galaxies. This implies that these high star forming galaxies will have their star formation quenched by $z \sim 0.6-0.8$ and will evolve into ellipticals. This also supports the assertion of downsizing because these highly star forming galaxies in dense environments will have turned into passive ellipticals leaving only star forming galaxies in less dense regions to the present day.

5.6 Conclusion

We conclude that environmental effects play a major role in the evolution of galaxies. We showed that high redshift ($z = 1.4$) galaxies selected on SFR were more clustered than spiral galaxies of the same stellar mass. Therefore we infer galaxies are star forming because of their environment at these redshifts and that stellar mass does not play an important role. Environmental effects produce gas instabilities which trigger star formation as opposed to the gas mass being triggered at a point in the galaxies life almost independently of the environment. We found downsizing was occurring by showing the clustering of SFR selected galaxies decreased toward $z = 0$, a trend which we found was mirrored in decreasing halo mass. The evolution of the halo clustering indicated high z SFR galaxies will evolve into ellipticals by $z \sim 0.5$. This shows it is the galaxies in the densest environments which have their star formation quenched leaving only those star forming galaxies in less dense regions by the present day, hence producing the observed downsizing. This is also supported by the constant relative bias with redshift (at least at $0.6 \leq z \leq 1.0$) which implied that it is the most clustered spirals, from the densest regions, which are turning into ellipticals to preserve the constant relative bias which we

expect to decrease. Therefore, environmentally induced effects are also responsible for turning off star formation, such as galaxy-galaxy interactions which possibly trigger AGN activity (Hopkins et al.(2007)), or processes such as harassment (Moore et al.(1999)) or ram pressure stripping (Gunn & Gott(1972)) which could be responsible for disrupting gas cooling. Therefore environment appears to play an important role in turning star formation on and off. These findings coherently show how important galaxy environment is to the evolution of galaxies. It is beyond the scope of this work to probe in detail which environmentally induced effects trigger and quench star formation. This would require a study tailored for such purposes.

Summary of key conclusions:-

- We found the colour-density relation was in place at $z = 1$. Therefore the mechanism(s) responsible for this relation was in place at $z > 1$. We found it to be constant at $0.6 \leq z \leq 1.0$ implying spiral galaxies in the densest regions are having their star formation quenched.

- The clustering strength of elliptical and spiral galaxies was invariant with stellar mass in the range $\log(M_{\star}/\mathrm{M}_{\odot}) \leq 11.2$ but increased with stellar mass at $\log(M_{\star}/\mathrm{M}_{\odot}) > 11.2$. We attribute this to more massive galaxies being more biased tracers of the dark matter distribution. We make the first detection of clustering of high mass galaxies ($\log(M_{\star}/\mathrm{M}_{\odot}) > 11.2$) at $z > 0.4$.

- We found the clustering strength of low mass spiral galaxies to decrease with increasing redshift indicating they have become more biased tracers of the dark matter distribution from $z = 1$ to $z = 0$. The clustering strength of ellptical galaxies showed less change over time. We conclude that structure formation is ongoing is less dense regions at later times and that denser regions of the Universe are largely in place by $z \sim 1$.

- Galaxies selected on SFR showed no variation in clustering strength at $\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) \leq 1.37$ at $z = 0.3$. At $z = 0.7$ from comparisons with Gilli et al.(2007) and Magliocchetti et al.(2008) there is an indication that clustering strength increases with SFR at $\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) > 2$. This is an area of research which requires further investigation.

- We made a clear detection of downsizing as the clustering strength of galaxies selected on SFR decreased strongly from $z = 1.4$ to $z = 0.3$. Although the high redshift galaxies had larger SFR we believe the strength of the decrease indicates this is the result of more than just the difference in SFR.

- At high redshift we found galaxies selected on SFR to be more strongly clustered than spiral galaxies of the same stellar mass. We conclude that environment plays an important role in making galaxies star forming and stellar mass is not important.

- Using the evolution of dark matter haloes as evolutionary tracks implied $\log(SFR/\mathrm{M}_{\odot}\mathrm{yr}^{-1}) = 2.85$ galaxies at $z = 1.4$ will evolve into high mass ellipticals by $z \sim 0.6$ showing it is the star forming galaxies in the densest environments which will have their star formation quenched.

- We conclude that environment is an important factor in turning star formation on at high redshift, more so than stellar mass. It also plays an important role in quenching star formation in galaxies.

5.7 Preliminary Analyses

The main clustering analysis presented in this thesis is detailed in Sections 5.2 and 5.3. Prior to the main analysis being conducted we performed some preliminary analyses using the photometric redshift catalogue whereby the clustering of *all* galaxies selected on stellar mass (i.e. not separated into ellipticals, spirals and starbursts) was measured. These analyses were conducted before some of the techniques used in the main analysis had been developed. Details of two of these analyses are given below as they highlight the necessity of using the full method.

5.7.1 First Preliminary Analysis

We split the galaxies into stellar mass-redshift cells in a similar manner to the main analysis however the actual cells used were not the same. Figure 5.29 shows the stellar mass-redshift cells used. This analysis also used the $r \leq 23.2$, $S_{3.6\mu m} > 10\mu$Jy and the 3.6μm binary mask. The method to reduce the number of photometric redshift outliers by defining n_{band} and χ^2 limits for each cell (as detailed in Section 3.7.4) had not been developed when this analysis was performed. Therefore no n_{band} and χ^2 limits were applied and

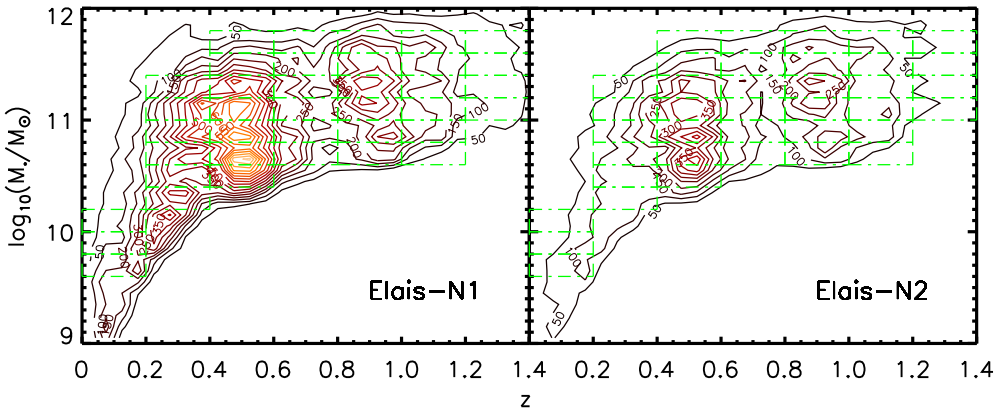


Figure 5.29: The number density of sources in the $M_\star - z$ plane for the first preliminary analysis indicated by the contours. The green squares define the $(M_\star, z)$ cells.

therefore good and bad photometric redshifts were used. The angular correlation function was measured using the same method used for the main analysis, as outlined in Section 5.2.6, with one difference. As no n_{band} and χ^2 limits were defined for each cell (and the technique for modeling this selection function (Section 5.2.4) had not yet been developed) no density map was applied to the random catalogue used in the correlation function pair counting. The comoving correlation strength r_0 was calculated in the same way using Limber's equation. Also at the time this analysis was carried out the stellar mass limits introduced by the r magnitude and $S_{3.6\mu m}$ cuts had not been calculated therefore cells at $z > 0.5$ where r_{cut} was biassed against certain types of galaxies could not be identified and hence were not removed from the analysis

In summary, this analysis differed from the main analysis in the following ways. 1) Different stellar mass-redshift cells were used. 2) Galaxies were not split into ellipticals, spirals and starbursts. 3) No n_{band} and χ^2 limits were used and hence photometric redshifts of good and bad quality were included. 4) Cells were not removed which were affected by the biasing of r_{cut}.

The comoving clustering strengths are shown in Figure 5.30. As the galaxies were not separated by type, the clustering strengths were a mixture of the clustering signal from both elliptical and spiral galaxies which cluster very differently. As there were more spiral than elliptical galaxies at any given redshift the combined clustering signal was weighted

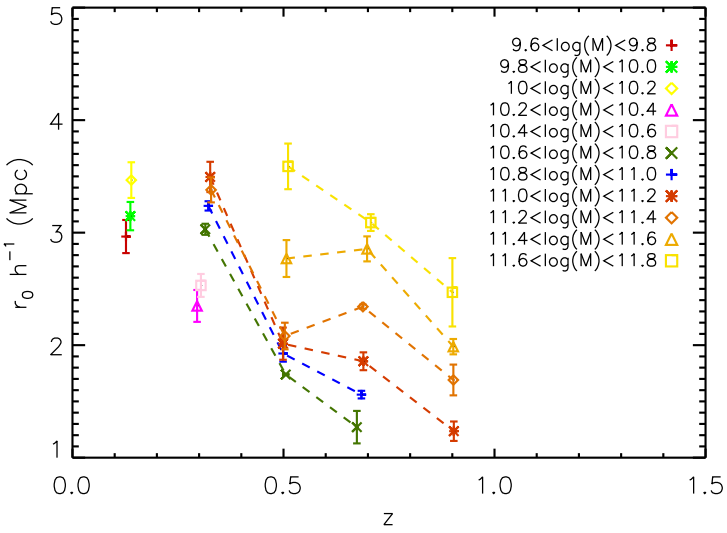


Figure 5.30: The comoving clustering strength r_0 as a function of redshift and $M_\star$ for the first preliminary analysis. Galaxies were not split by type.

more toward that of the spiral galaxies leading to a depressed r_0 when compared with the typical values in Figure 5.21.

More importantly a substantial fraction of the photometric redshifts used here are very inaccurate as no constraints on n_{band} and χ^2 were used. The inaccurate photometric redshifts used suppressed the clustering signal. This is most clear to see at $z = 0.5$ where the redshift distribution peaks (Figure 3.9), r_0 dips for a range of stellar masses. We attribute this to the fact that the peak will contain more inaccurate photometric redshifts. These results should not be trusted and confirms the need to use only accurate photometric redshifts.

5.7.2 Second Preliminary Analysis

The second preliminary analysis also measured the clustering of *all* galaxy types selected on stellar mass. This analysis used the same method as the first preliminary analysis (Section 5.7.1) except only good photometric redshifts were used by applying n_{band} and χ^2 cuts for each cell. Density maps were made for each cell following the method in Section 5.2.4 and were applied to the random catalogues used in the pair counting. The stellar mass-redshift cells are shown in Figure 5.31.

The comoving clustering strengths are shown in Figure 5.32. r_0 was a mixture of the clustering signal from ellipticals and spirals, again weighted to lower r_0 values by the

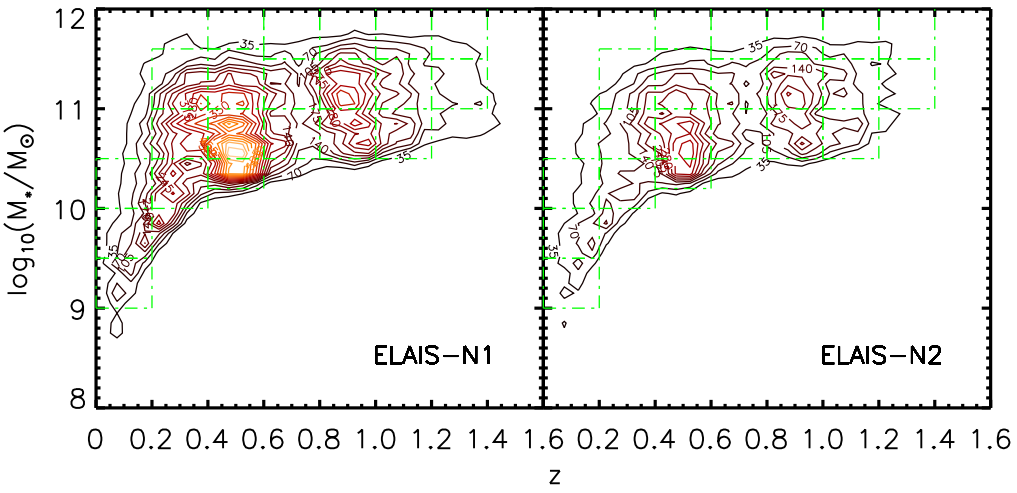


Figure 5.31: The number density of sources in the $M_\star - z$ plane, for the second preliminary analysis, as indicated by the contours. The green squares define the $(M_\star, z)$ cells.

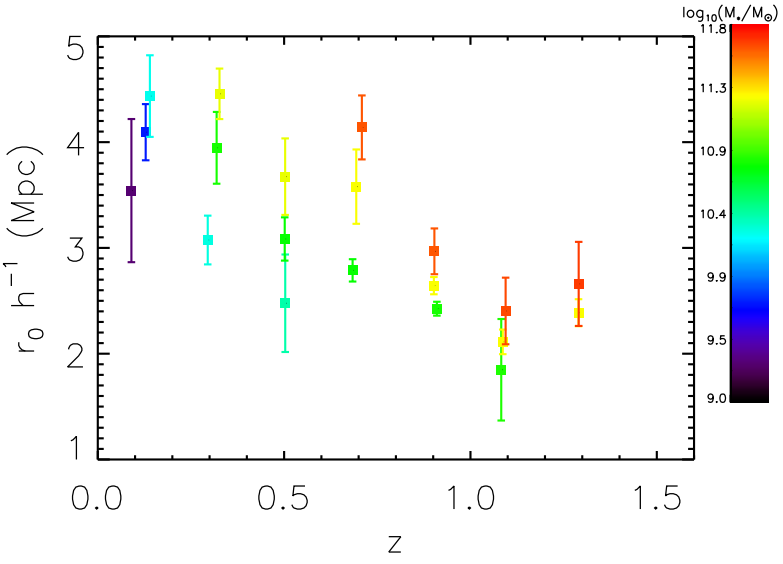


Figure 5.32: The comoving clustering strength r_0 as a function of redshift for the second preliminary analysis. Each data point is coloured to represent the $M_\star$ of the cell indicated by the colour scale on the right hand side.

number of spirals. The dip at $z = 0.5$ has been removed by using good photometric redshifts. A steady increase in clustering strength with stellar mass is seen at all redshifts however the ratio of elliptical to spiral galaxies will have also increased with stellar mass.

5.8 Summary

In this chapter we constrained the selection effects, in particular incompleteness, discussed in Chapter 3 by applying flux and coverage cuts and the masks. This was essential to ensure a homogeneous data set was used, suitable for a clustering analysis. n_{band} and χ^2 cuts were applied to each $(M_\star, z)$ and (SFR, z) cell to ensure only good photometric redshifts were used. The removal of galaxies due to these cuts was modeled by selecting galaxies from the across the field with similar observable properties to those in the cell in question. Such galaxies would respond to the n_{band} and χ^2 cuts in the same manner as the galaxies in the cell. The n_{band} and χ^2 cuts were applied to these galaxies and mapped as a function of position which was then applied to the random catalogue used in the angular correlation function pair counting.

We measured the clustering of elliptical, spiral and starburst galaxies selected on stellar mass, and spiral galaxies selected on SFR, over $0.1 < z < 1.5$. We ascertained typical parent halo masses by comparing the clustering strengths of SWIRE galaxies to those of Millennium Simulation haloes. Our key results are summarised in Section 5.6. We conclude from our results that galaxies are star forming because of their environment at high redshifts and that stellar mass does not play an important role. We also conclude it is the spiral galaxies in the densest environments which have their star formation quenched to turn into passive elliptical galaxies.

5.9 Tables of Data

Table 5.4: This table shows information for elliptical galaxies in each $(M_\star, z)$ cell. The information in each column is as follows; the cell number, mean redshift, mean stellar mass, number of galaxies N_{gal}, mean B-band luminosity and the spatial correlation length, r_0.

Cell	z	$\log(M_\star/\mathrm{M}_\odot)$	N_{gal}	$\log(L_B/\mathrm{L}_\odot)$	r_0 h^{-1} Mpc
M0	0.21	10.26	680	9.88	5.05 ± 0.12
M1	0.22	10.77	1166	10.38	4.98 ± 0.16
M2	0.23	11.21	1047	10.81	5.05 ± 0.50
M3	0.37	10.39	202	9.98	-
M4	0.40	10.78	2010	10.35	4.85 ± 0.14
M5	0.40	11.23	2216	10.85	4.77 ± 0.11
M6	0.39	11.77	315	11.38	6.31 ± 0.58
M7	0.55	10.86	630	10.49	-
M8	0.59	11.22	2019	10.88	4.77 ± 0.38
M9	0.61	11.62	301	11.33	5.22 ± 0.41
M10	0.71	10.96	6	10.76	-
M11	0.76	11.34	203	11.08	-
M12	0.79	11.65	258	11.40	4.93 ± 0.42
M13	-	-	-	-	-
M14	0.93	11.41	6	11.23	-
M15	0.97	11.68	267	11.43	4.45 ± 0.95
M16	-	-	-	-	-
M17	-	-	-	-	-
M18	1.14	11.94	7	11.76	-
M19	-	-	-	-	-
M20	-	-	-	-	-

Table 5.5: This table shows halo information for elliptical galaxies in $(M_\star, z)$ cells. The information in each column is as follows; mean redshift, mean stellar mass, parent halo mass, M_{hz}, halo mass at redshift zero, M_{h0} and comoving spatial clustering strength, r_0. Note, low mass cells ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) at each redshift from Table 5.4 were averaged together producing cells EM0-EM2 shown in this Table.

Cell	z	$\log(M_\star/\mathrm{M}_\odot)$	$\log(M_{hz}/\mathrm{M}_\odot)$	$\log(M_{h0}/\mathrm{M}_\odot)$	r_0 h^{-1} Mpc
EM0	0.2	≤ 11.2	$12.3^{+0.1}_{-0.1}$	$12.7^{+0.1}_{-0.1}$	5.03 ± 0.09
EM1	0.4	≤ 11.2	$12.3^{+0.1}_{-0.1}$	$12.8^{+0.1}_{-0.1}$	4.83 ± 0.09
EM2	0.6	≤ 11.2	$12.3^{+0.2}_{-0.3}$	$12.9^{+0.1}_{-0.2}$	4.78 ± 0.38
EM3	0.6	11.6	$12.6^{+0.1}_{-0.3}$	$13.1^{+0.2}_{-0.1}$	5.22 ± 0.41
EM4	0.8	11.6	$12.4^{+0.2}_{-0.2}$	$13.1^{+0.2}_{-0.1}$	4.93 ± 0.42
EM5	1.0	11.6	$12.2^{+0.3}_{-0.9}$	$13.0^{+0.3}_{-0.4}$	4.45 ± 0.95

Table 5.6: This table shows information for spiral galaxies in each $(M_\star, z)$ cell. The information in each column is as follows; the cell number, mean redshift, mean stellar mass, number of galaxies N_{gal}, mean B-band luminosity and the spatial correlation length, r_0.

Cell	z	$\log(M_\star/\mathrm{M}_\odot)$	N_{gal}	$\log(L_B/\mathrm{L}_\odot)$	r_0 h^{-1} Mpc
M0	0.22	10.25	2363	10.09	-
M1	0.24	10.73	1837	10.49	4.34 ± 0.19
M2	0.24	11.21	808	10.87	4.28 ± 0.36
M3	0.42	10.38	2665	10.26	-
M4	0.41	10.76	7174	10.51	3.49 ± 0.71
M5	0.41	11.20	3988	10.91	3.68 ± 0.12
M6	0.41	11.70	194	11.36	-
M7	0.58	10.75	5458	10.65	3.14 ± 0.07
M8	0.58	11.20	3543	10.96	3.15 ± 0.58
M9	0.58	11.60	243	11.31	-
M10	0.80	10.85	2884	10.77	-
M11	0.83	11.23	2548	11.10	2.33 ± 0.27
M12	0.84	11.61	490	11.41	3.29 ± 0.35
M13	0.95	10.93	2751	10.97	-
M14	0.98	11.26	2753	11.16	2.25 ± 0.19
M15	1.00	11.63	1215	11.44	2.71 ± 0.47
M16	1.14	10.95	1113	11.11	-
M17	1.18	11.32	1742	11.28	2.02 ± 0.24
M18	1.19	11.66	736	11.51	2.80 ± 0.19
M19	1.37	11.31	1043	11.38	-
M20	1.40	11.65	683	11.61	2.42 ± 0.45

Table 5.7: This table shows halo information for spiral galaxies in $(M_\star, z)$ cells. The information in each column is as follows; mean redshift, mean stellar mass, parent halo mass, M_{hz}, halo mass at redshift zero, M_{h0} and comoving spatial clustering strength, r_0. Note, low mass cells ($\log(M_\star/\mathrm{M}_\odot) \leq 11.2$) at each redshift from Table 5.6 were averaged together producing cells SPM0-SPM5 shown in this Table.

Cell	z	$\log(M_\star/\mathrm{M}_\odot)$	$\log(M_{hz}/\mathrm{M}_\odot)$	$\log(M_{h0}/\mathrm{M}_\odot)$	r_0 h^{-1} Mpc
SPM0	0.2	≤ 11.2	$11.7^{+0.2}_{-0.2}$	$12.2^{+0.1}_{-0.1}$	4.31 ± 0.17
SPM1	0.4	≤ 11.2	$11.0^{+0.1}_{-0.2}$	$11.9^{+0.1}_{-0.3}$	3.59 ± 0.12
SPM2	0.6	≤ 11.2	$10.6^{+0.1}_{-0.1}$	$11.7^{+0.1}_{-0.2}$	3.15 ± 0.08
SPM3	0.8	≤ 11.2	$10.1^{+0.2}_{-0.2}$	$11.1^{+0.1}_{-0.1}$	2.34 ± 0.27
SPM4	1.0	≤ 11.2	$10.0^{+0.2}_{-0.1}$	$11.4^{+0.1}_{-0.3}$	2.25 ± 0.19
SPM5	1.2	≤ 11.2	$10.0^{+0.2}_{-0.2}$	$11.5^{+0.1}_{-0.4}$	2.02 ± 0.24
SPM6	0.8	11.6	$10.9^{+0.6}_{-0.4}$	$12.4^{+0.2}_{-0.3}$	3.29 ± 0.35
SPM7	1.0	11.6	$10.5^{+0.4}_{-0.5}$	$12.1^{+0.4}_{-0.5}$	2.71 ± 0.47
SPM8	1.2	11.6	$10.7^{+0.1}_{-0.2}$	$12.4^{+0.2}_{-0.1}$	2.80 ± 0.19
SPM9	1.4	11.6	$10.4^{+0.4}_{-0.4}$	$12.3^{+0.2}_{-0.5}$	2.42 ± 0.45

Table 5.8: This table shows information for starburst galaxies in each $(M_\star, z)$ cell. The information in each column is as follows; the cell number, mean redshift, mean stellar mass, number of galaxies N_{gal}, mean B-band luminosity and the spatial correlation length, r_0.

Cell	z	$\log(M_\star/\mathrm{M}_\odot)$	N_{gal}	$\log(L_B/\mathrm{L}_\odot)$	r_0 h^{-1} Mpc
M0	0.24	10.21	554	10.37	2.36 ± 0.32
M1	0.25	10.69	301	10.84	2.95 ± 0.73
M2	0.24	11.18	39	11.31	-
M3	0.43	10.35	368	10.51	3.15 ± 0.38
M4	0.43	10.71	371	10.83	3.16 ± 0.72
M5	0.42	11.16	58	11.30	3.49 ± 0.07
M6	0.00	0.00	0	-	-
M7	0.58	10.69	1305	10.86	2.89 ± 0.14
M8	0.58	11.17	394	11.32	3.91 ± 0.24
M9	0.60	11.53	9	11.62	-
M10	0.82	10.71	379	10.88	2.32 ± 0.05
M11	0.84	11.18	40	11.29	2.85 ± 0.40
M12	0.84	11.58	3	11.66	-
M13	0.99	10.76	562	11.01	2.32 ± 0.37
M14	0.99	11.16	319	11.33	2.56 ± 0.17
M15	1.01	11.60	42	11.72	-
M16	1.18	10.85	341	11.12	-
M17	1.19	11.17	333	11.38	2.52 ± 0.58
M18	1.20	11.60	35	11.74	-
M19	1.47	11.12	244	11.42	2.92 ± 029
M20	1.57	11.58	81	11.79	-

Table 5.9: This table shows halo information for starburst galaxies in $(M_\star, z)$ cells. The information in each column is as follows; mean redshift, mean stellar mass, parent halo mass, M_{hz}, halo mass at redshift zero, M_{h0} and comoving spatial clustering strength, r_0. Note, low mass cells were not averaged together as they were for spiral and elliptical galaxies. Therefore these cells are the same as those in Table 5.8.

Cell	z	$\log(M_\star/\mathrm{M}_\odot)$	$\log(M_{hz}/\mathrm{M}_\odot)$	$\log(M_{h0}/\mathrm{M}_\odot)$	r_0 h^{-1} Mpc
SBM0	0.2	10.7	$10.1^{+0.7}_{-0.6}$	$9.9^{+1.4}_{-0.3}$	2.95 ± 0.73
SBM1	0.4	10.7	$10.5^{+0.8}_{-0.8}$	$11.2^{+1.0}_{-0.3}$	3.16 ± 0.72
SBM2	0.6	10.7	$10.4^{+0.1}_{-0.2}$	$11.2^{+0.2}_{-0.2}$	2.89 ± 0.14
SBM3	0.8	10.7	$10.0^{+0.1}_{-0.1}$	$11.0^{+0.1}_{-0.1}$	2.32 ± 0.05
SBM4	1.0	10.7	$10.1^{+0.4}_{-0.3}$	$11.6^{+0.5}_{-0.6}$	2.32 ± 0.37
SBM5	0.4	11.2	$10.9^{+0.1}_{-0.1}$	$11.7^{+0.2}_{-0.2}$	3.49 ± 0.07
SBM6	0.6	11.2	$11.5^{+0.3}_{-0.3}$	$12.5^{+0.1}_{-0.2}$	3.91 ± 0.24
SBM7	0.8	11.2	$10.5^{+0.4}_{-0.3}$	$12.1^{+0.2}_{-0.8}$	2.85 ± 0.40
SBM8	1.0	11.2	$10.3^{+0.2}_{-0.1}$	$12.0^{+0.2}_{-0.3}$	2.56 ± 0.17
SBM9	1.2	11.2	$10.4^{+0.5}_{-0.5}$	$12.2^{+0.4}_{-0.8}$	2.52 ± 0.58
SBM10	1.4	11.2	$10.8^{+0.3}_{-0.2}$	$12.7^{+0.1}_{-0.1}$	2.92 ± 0.29

Table 5.10: This table shows information for each (SFR, z) cell. The information in each column is as follows; mean redshift, number of galaxies N_{gal}, mean $\log(L_{IR}/\mathrm{L}_\odot)$, ratio of IR luminosity to a $L_{IR} = 10^{11}\mathrm{L}_\odot$ galaxy, mean SFR, mean stellar mass, typical parent halo mass and the spatial correlation length, r_0.

Cell	z	N_{gal}	$\log(L_{IR}/\mathrm{L}_\odot)$	L_{IR}/L^*	$\log(SFR/\mathrm{M}_\odot\mathrm{yr}^{-1})$	$\log(M_\star/\mathrm{M}_\odot)$	$\log(M_{hz}/\mathrm{M}_\odot)$	r_0 $h^{-1}\mathrm{Mpc}$
S0	0.27	1161	10.53	0.96	0.68	10.94	$11.0^{+0.1}_{-0.2}$	3.18 ± 0.31
S1	0.31	1173	11.12	1.01	1.37	11.14	$11.0^{+0.5}_{-0.6}$	3.16 ± 0.46
S2	0.71	998	11.59	1.05	1.86	11.34	$12.8^{+0.5}_{-0.9}$	4.38 ± 1.39
S3	1.06	1530	12.01	1.09	2.29	11.48	$12.4^{+0.6}_{-1.4}$	2.90 ± 1.26
S4	1.38	116	12.56	1.14	2.85	11.59	$13.1^{+0.3}_{-0.3}$	4.24 ± 0.85

Chapter 6

A Search for Population III Stars

6.1 Introduction

In this chapter a search for population III stars is undertaken by trying to detect the supernovae which mark the end of their lives. The search is conducted in the Spitzer IRAC dark field. The dark field has repeat imaging on intervals of 2–3 weeks over a 2.2 year baseline. The individual epochs reach a typical $5\,\sigma$ depth of $1\,\mu$Jy at 3.6μm. The unprecedented depth and multi-epochal nature of the data make it ideal for a first foray to try to detect transient objects which may be candidate luminous pair instability supernovae from the primordial-metallicity first stars. Models of population III stars from the literature are used as a guide for typical luminosities, redshifts and light curves. The search is conducted in catalogue space. Spitzer imaging data, as well as deep *HST*/ACS F814W imaging available in the field, are used to identify potential candidates.

6.2 Population III Stars

Primordial metallicity Population III stars are thought to be the first luminous objects to form in the Universe. Their formation marks the end of the cosmic dark ages and an important transition of the Universe from a homogeneous state to a highly structured one. The UV photons produced by such stars at high redshifts are also thought to be at least partly responsible for re-ionizing the universe (Tumlinson & Shull(2000), Bromm et al.(2001), Schaerer(2002), Schaerer(2003)). It is believed that the explosive events that mark the end state of such stars seed the intergalactic medium with heavy elements

(Gnedin & Ostriker(1997), Furlanetto & Loeb(2003), Greif et al.(2007)). Hence, studying these objects is of great importance in helping us to understand the high-redshift universe.

To date, the supernovae marking the deaths of first stars (Pop III SNe) at high redshift have not been observed (though see Woosley et al.(2007)). Theoretical models provide a better understanding of the properties of such stars and when they might have existed. It is thought that Population III stars formed out of primordial-abundance H/He gas in low-mass dark matter haloes. For primordial-abundance stars it is expected that the explosion mechanism may drive not only "classical" supernovae, but also "hypernovae" for certain progenitor masses, driven through an electron-positron instability mechanism that results in explosive events with up to approximately one hundred times greater luminosities (Umeda & Nomoto(2002)). Exactly when the epoch of the first stars began is still a matter of debate but estimates place it at $10 < z < 50$ (Wise & Abel(2005)). Thereafter, such objects could exist in primordial-metallicity pockets even at relatively low redshifts, even $z \leq 2.5$ (e.g. Scannapieco(2005), Tornatore et al.(2007)). The primordial metallicity of Pop III stars leads to inefficient cooling mechanisms through H_2, leading to very high stellar masses and to a top-heavy initial mass function (IMF) with a large fraction of stars having $M_\star > 100 M_\odot$ (Bromm et al.(1999), Bromm et al.(2002), Abel et al.(2000), Abel et al.(2002)) and possibly even greater masses (Bromm & Loeb(2004), Omukai & Palla(2003)).

6.3 The IRAC dark field

The search for Pop III SNe was conducted in the Spitzer IRAC dark field. The IRAC dark field is the dark current calibration target for IRAC (Fazio et al.(2004)) on board Spitzer. It is an extragalactic field of very low background, in Spitzer's continuous viewing zone near the North Ecliptic Pole. This area was observed at the start and end of each IRAC observing campaign (2–3 weeks apart) since Spitzer's first light. For technical reasons anchored on the need for high-quality dark frames and on the normal precession of the observatory, there is only a modest overlap in the observed area on the sky from epoch to epoch.

The data we used in this analysis was based on 128 distinct epochs over the first 2.2 years of Spitzer's operations. Each epoch was composed of multiple individual exposures at all of the available IRAC exposure times. The full IRAC mosaic had a total observing time

of > 500 hours, and covered an approximately circular area, 20 arcminutes in diameter. Each point in the mosaic typically had more than 10 hours of total integration time, with a maximum of ~ 80 hours in the area of maximal overlap across epochs. It was classically confusion-limited in all four IRAC channels (3.6μm, 4.5μm, 5.8μm, 8.0μm). The IRAC source extraction was done using Source EXtractor, Bertin & Arnouts(1996).

The unprecedented depth and multi-epochal nature of the data make it ideal for a first foray into trying to detect supernovae from the first stars. The field will, in fact, continue to be observed through the full extent of the Spitzer Warm Mission, eventually giving a baseline of around seven (and possibly ten) years. There is now a wealth of multiwavelength data available in the dark field including *Chandra* and *HST*/ACS F814W imaging, Krick et al.(2008). Figure 6.1 shows the dark field 3.6μm imagery with the HST/ACS imagery overlaid in blue.

6.3.1 Epochal Database

A common method used to find transient objects, is to search for significant residuals when differencing registered time-series observations. However, because the IRAC point spread function is asymmetric and fairly complex, and indeed had a different orientation on the image of each epoch, it is difficult to distinguish candidate transient objects from artifacts in the difference images. Therefore, our search was conducted through cross-correlating catalogues of objects detected in each individual epoch.

Each epoch of observation had a typical coverage (a coverage of 1 corresponds to 100 second exposure) of $4-10$ or 3.6μm 5σ depth of 1 μJy. Figure 6.2 shows the $5\,\sigma$ sensitivity as a function of coverage, calculated using the Sensitivity Performance Estimation Tool SENSPET[1]. SENSPET calculates the sensitivity in μJy based on coverage, number of repeat images and background level. The corresponding $m_{3.6\mu\mathrm{m}}$(AB) are shown in Figure 6.3. The mean coverage in the field corresponds to $m_{3.6\mu\mathrm{m}}(\mathrm{AB}) \sim 23.7$.

We collated the epochal catalogues together by cross-matching them using a $1''$ radius which gave a database of 31,492 sources, each of which had at least one detection in a single distinct epoch. The dark field had good astrometry from epoch to epoch shown by the small drift in position, see Figure 6.4, which made cross matching the catalogues easy.

The master database contained all the critical information for a transient search. For each object, we recorded the following: in which epochs the object was within the observed

[1]SENSPET is an online tool found at http://ssc.spitzer.caltech.edu/tools/senspet/

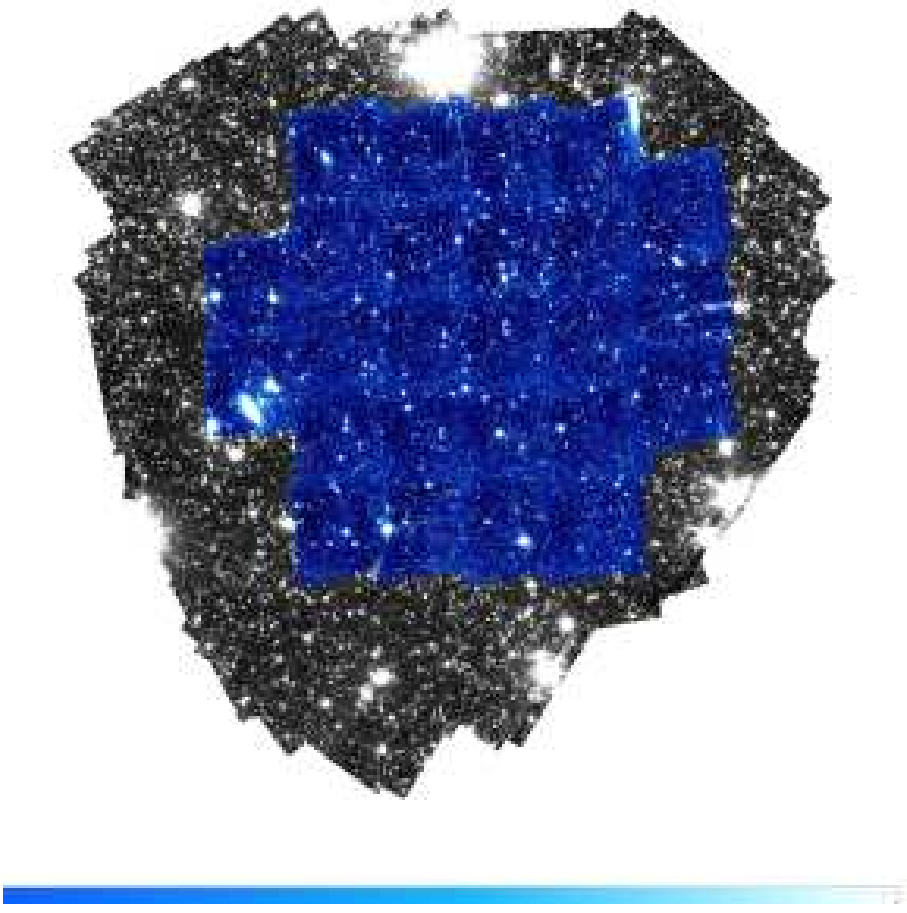


Figure 6.1: The IRAC dark field. The 3.6μm imagery is shown in black with the ACS HST imagery overlaid in blue.

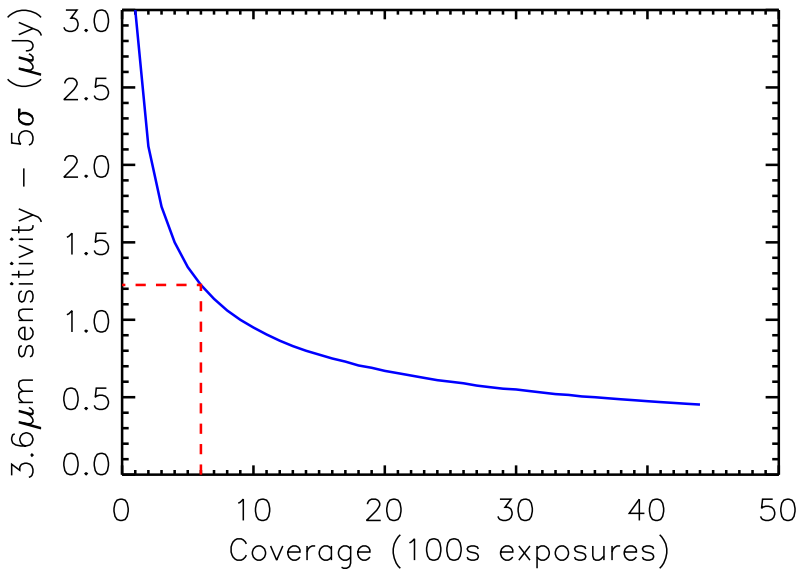


Figure 6.2: IRAC $5\,\sigma$ sensitivity in the dark field as a function of coverage. The dotted red line indicates the mean epochal coverage and corresponding sensitivity.

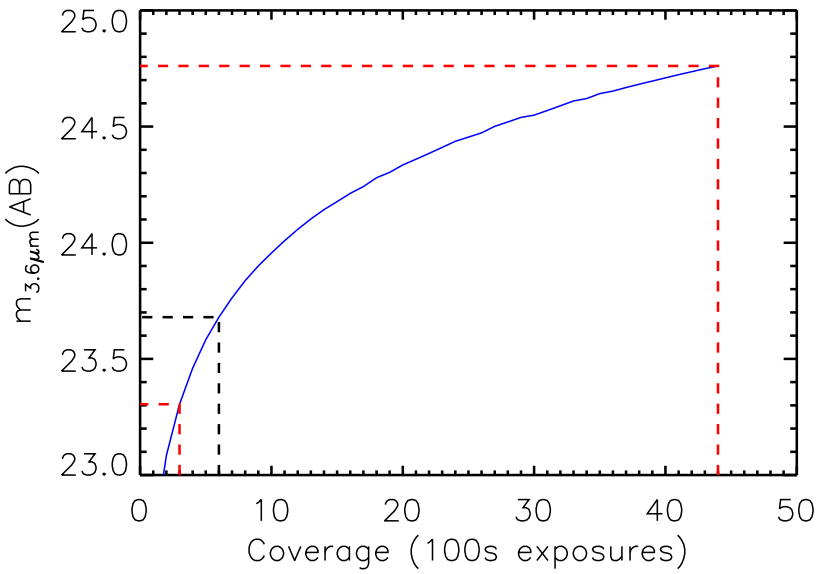
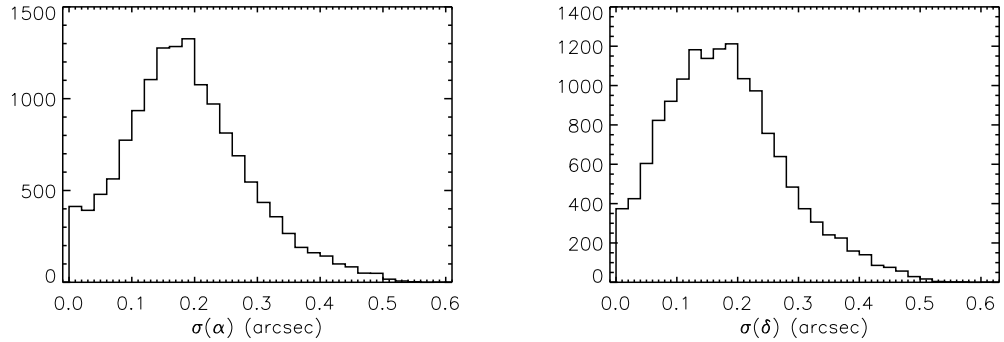


Figure 6.3: The $m_{3.6\mu\mathrm{m}}$(AB) corresponding to the sensitivity limits as a function of coverage in the dark field. The black dotted line indicates the mean coverage in the field and corresponding $m_{3.6\mu\mathrm{m}}$(AB). Red dotted lines show the maximum and minimum coverages.



(a) α (b) δ

Figure 6.4: Histograms showing the typical drift in position between epochs in a) right ascension and b) declination. The drift was negligible shown by the modal value of ∼0.2 arcsec.

field and detected, and its flux density; in which epochs the object was within the observed field and *not* detected, and the known sensitivity at that position (based on the coverage). If the spectral energy distributions of high-redshift Pop III SNe were above the detection limits in some epochs within the timespan of this survey, they would appear as transient objects.

6.4 Expectations from Theory

Massive Pop III stars with primordial metallicity are thought to be common at high redshift. Stars with $M_\star < 140\mathrm{M}_\odot$ or $M_\star > 260\mathrm{M}_\odot$ are thought to form black holes at the end of their evolution (e.g. Fryer et al.(2001), Heger & Woosley(2002)). Those which have masses between 140–260 $M_\odot$ are thought to end their lives as pair-instability supernova (PISNe). Once helium burning in the core of such stars has ceased there is sufficient entropy to create positron-electron pairs Wise & Abel(2005). This process converts thermal energy to the mass of the particle pair and the pressure in the core is reduced. This leads to a partial collapse which triggers a thermonuclear explosion. The star is completely destroyed leading to a PISN, in which no remnant is left behind. At least one relatively local analog may already have been observed (SN 2006gy in NGC 1260; Smith et al.(2007), Woosley et al.(2007)), which lends support to the possibility of this mechanism. PISNe would be "host-less" and as much as one hundred times more luminous

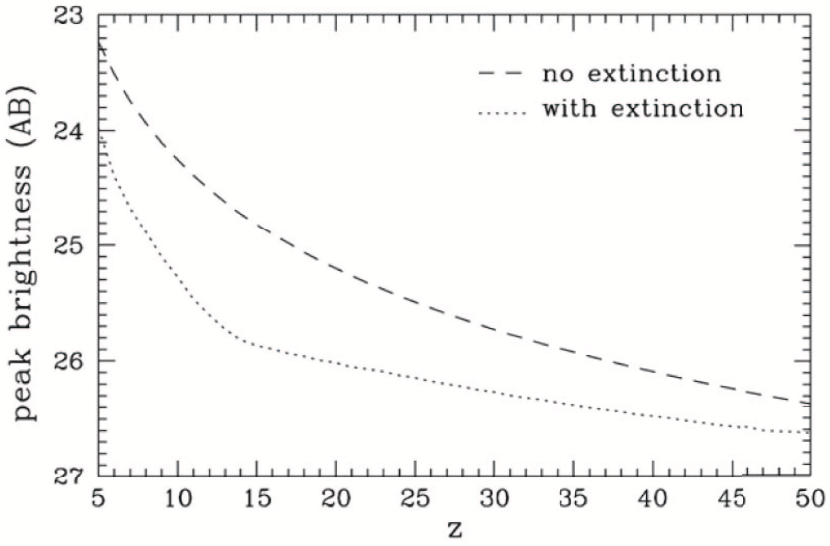


Figure 6.5: The peak AB apparent magnitude of a 250$M_\odot$ PISNe as observed at the current epoch when the PISNe detonates at a range of redshifts, Weinmann et. al. 2004. The dotted curve shows the worst case extinction and the dashed curve with no extinction.

than more-typical supernovae.

The best chance of detecting a pop III SNe is around the peak in it's light curve. Precisely how far down the light curve that we could have detected depends on the luminosity of the SNe, it's redshift, the rate of decay of the light curve and the depth of the dark field at that location and time. The only known quantity is the depth of the dark field. Since no direct observations of pop III SNe exist, we used predictions from the literature to anticipate luminosities, durations (through their light curves), and frequency of events. Scannapieco(2005), using light curves calculated by Weinmann & Lilly(2005), predict peak apparent magnitudes of $m_{\rm AB} \sim 26.8$ at $z = 10$ for $250\,{\rm M}_\odot$ PISN, assuming negligible extinction. This suggests that the typical by-epoch depth of $m_{3.6\mu{\rm m}}({\rm AB}) \sim 24$ of our search may be able to detect such objects at $z \sim 3-5$.

PISN light curves were calculated in Wise & Abel(2005), Heger et al.(2002), and Scannapieco(2005). A broad plateau phase is expected, which could last from ∼10 days to as long as a full year in the frame of the event. Since in the observed frame the light curve is stretched by a factor of $(1+z)$, there could be events that would appear as near-continuous sources over the entire 2.2-year baseline of the current dataset. Ultimately the dataset could span 7- to 10-years by the end of the *Spitzer* Warm Mission. Therefore there is

Table 6.1: Predicted Population III supernova rates found in the literature. The rate of Heger et. al. 2002 and Cen et. al. 2003 have taken into account the correction factors of $(1+z)^{-2}$ and $(1+z)^{-1}$ determined by Weinmann et. al. 2004.

dN/dz ($\deg^{-2}\mathrm{yr}^{-1}$)	$M_{progenitor}$ ($M_\odot$)	z	Reference
∼0.2	250	20	Heger et al.(2002)
0.34	100-500	10	Wise & Abel(2005)
50	250	> 15	Mackey et al.(2003)
11	100	> 13	Cen(2003)
25	140-260	5	Weinmann & Lilly(2005)

great potential in future analyses to encompass longer plateau-duration SNe.

6.4.1 Predicted Pop III SNe rates

Table 6.1 lists several predicted Pop III SNe differential rates from the literature. The rates quoted in Table 6.1 for Heger et al.(2002) and Cen(2003) incorporate the corrections determined by Weinmann & Lilly(2005) of $(1+z)^{-2}$ and $(1+z)^{-1}$, respectively. These rates are over redshift ranges beyond the sensitivity of our search, but are quoted here for completeness. Wise & Abel(2005) found a Pop III SNe rate of $0.34\,\deg^{-2}\mathrm{yr}^{-1}$ at $z = 10$, which changed negligibly over the mass range $100\mathrm{M}_\odot < \mathrm{M}_\star < 500\mathrm{M}_\odot$.

A wide range of values are expected, $\sim 0.2-50\,\deg^{-2}\ \mathrm{yr}^{-1}$, which is indicative of how the parameters involved in such predictions are still not well constrained. By the review in Weinmann & Lilly(2005), realistic rates are expected to be $dN/dz \sim 4\,\deg^{-2}\mathrm{yr}^{-1}$ for $z > 15$ and $0.2\,\deg^{-2}\mathrm{yr}^{-1}$ for $z > 25$. Also see Scannapieco(2005) for further discussion of these predictions.

6.5 Search Method

The search for Pop III SNe was conducted using the master database described in Section 6.3.1. Particular attention was paid toward transient objects. We conducted systematic searches for three different ranges of possible light curve durations (in the observed frame), which were appropriate to the 2.2-year baseline of the present dataset, and which were

plausible based on some of the theoretical expectations for the light curves. These were $0 < t_{obs} < 100$, $100 < t_{obs} < 200$, and $200 < t_{obs} < 400$ days. In addition to the practical convenience of analyzing the data in this compartmentalized fashion, it also anticipated the possibility that some transient-source light curves may have been truncated by either end of our 2.2 year baseline. Also it was not possible to know how far down a Pop III SNe light curve we could detect as this depended on the luminosity of the SNe, the decay of the light curve, the exact redshift at which it was located and the depth of the field at that point and time. Therefore the range of durations probed helped take into account these factors.

6.5.1 Foreground Stars

The precise pointing (and orientation) of Spitzer varied from epoch to epoch. Therefore this changed the on-sky orientation of the IRAC diffraction pattern. This essentially lead to rotating diffraction spikes across images. The photometry of objects near bright stars was systematically contaminated by the rotating diffraction spikes. We identified and removed the stars, and sources close enough to them to be affected, from the master catalogue as follows.

Imaging with *HST*/ACS F814W was obtained between November & December 2006, one year after the end of the IRAC dataset described here. Figure 6.6 shows a plot of the Source EXtractor (Bertin & Arnouts(1996)) ISOAREA versus aperture magnitude photometry. The dark field sources clearly separated into unresolved stars and resolved galaxies because unresolved sources have smaller ISOAREA than galaxies at any given magnitude. This method identified 1147 stars. We empirically determined a radius of $r = 0.06 \times S_{3.6\mu m}$ arcsec (but no greater than 30 arcsec) to mask out a circular area around each star which also removed an additional 5858 sources which were in close proximity. Figure 6.7 shows the star mask. Sources located within black areas were removed from the search.

6.5.2 Quality Criteria

Only part of the dark field was observed at each epoch, therefore the light curves of all sources were discontinuous as the field of view shifted. Also the depth of the field was variable from epoch to epoch and hence faint objects, such as pop III SNe, could potentially have appeared intermittent as they dropped out at shallow epochs. Therefore we defined

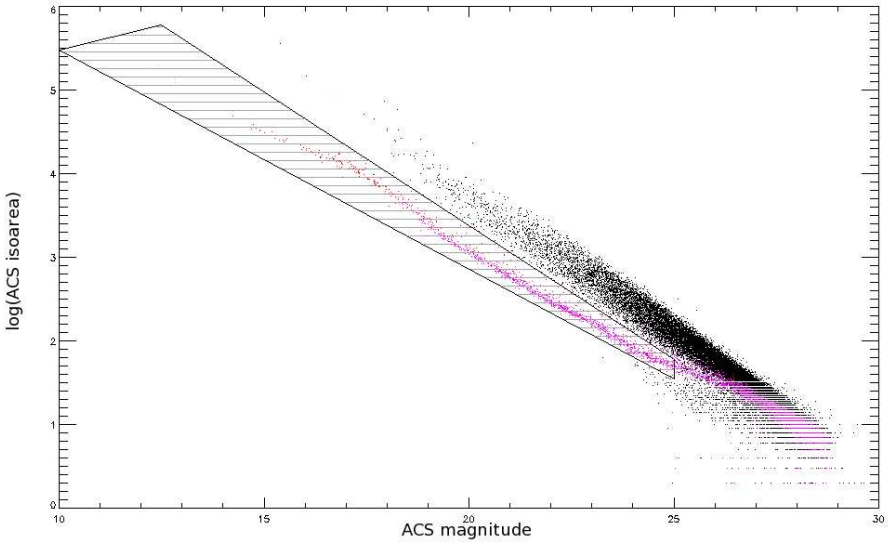


Figure 6.6: ACS isoarea against ACS magnitude for sources in the dark field. The plot separates into two branches with stars occupying the lower one (pink points) as stars always have a smaller isoarea than galaxies (black points) at any given magnitude.

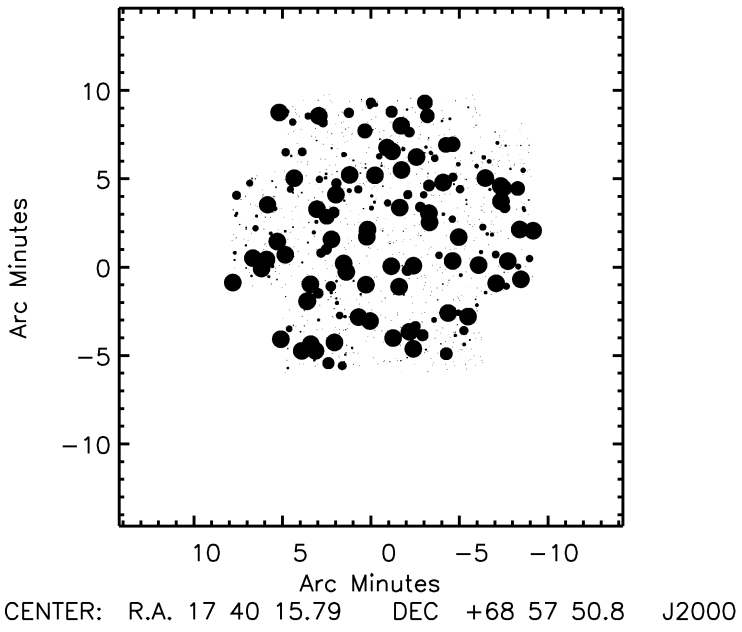


Figure 6.7: The star mask used to exclude stars and surrounding sources affected by their diffraction spikes which changed orientation with epoch. Black regions were excluded from the search and white region were included.

broad criteria to take into account these properties of the data and the uncertainties in the properties of the SNe indicated from simulations.

Given the relative non-homogeneity in the survey's depth and variations between epochs, we applied several criteria to remove erroneous sources. 1) An *upper* flux density limit of 40 μJy was imposed, which was an empirically determined practical threshold for removing additional objects that were affected by diffraction spike artifacts, beyond the masked radii, and object de-blending issues. 2) For all sources only epochs which had a 5 σ detection of >1 μJy were included. 3) Sources which only appeared in a *single* epoch were not considered. 4) Sources detected in only two or three epochs were only included in the search if they had 5 σ detections of >1 μJy in *all* epochs. 5) Any object with multiple epoch detections required the 3.6μm flux uncertainty to be less than 10% in at least half of its detected epochs. This ensured that a source was not rejected due to having large uncertainties in a small number of epochs (e.g. due to some epochs being particularly shallow relative to the others). 6) Finally, at least two significant non-detections were required, to ensure a clear transient signal. These criteria as a whole were very broad by design so there was no danger of rejecting a potential SNe.

After the criteria were applied 650 candidates remained that warranted more careful follow-up. We visually inspected the 3.6μm light curves and the corresponding IRAC and *HST* imaging for all 650 objects. No a priori restrictions were placed on the shape of the light curve. Several objects with otherwise flat light curves were found to have a dramatic "flare-up" in a single epoch. Careful inspection of the individual IRAC 100 s exposures of that epoch showed the spike to be a cosmic ray incident. The remaining candidates were found to have resolved-source counterparts in the *HST* data, and are therefore candidate low-redshift active galactic nuclei, which though extremely interesting in their own right, are clearly not Pop III SNe candidates. At the end of this careful analysis and vetting procedure, *no* viable candidates survived. In the next section the formal limits implied by the search are calculated.

6.6 Discussion

Despite the thorough search method adopted no Pop III SNe were detected over the baseline of 2.2 years to the sensitivity limit of $m_{3.6\mu\mathrm{m}}(\mathrm{AB}) \sim 24$ in the IRAC dark field. The rate of (non) detections can be calculated using,

$$R = \frac{n}{\Delta\Omega\Delta t} \tag{6.1}$$

where R is the rate of detection, $\Delta\Omega$ is the area of the field and Δt is the baseline. Assuming $n = 1$ and based on a usable area of 214 arcminutes^2 over the 2.2 year baseline, a rate of 8 deg^{-2} yr^{-1} is found. As no detections were actually made this rate is an upper limit. This can be taken a step further. The detection of supernovae can be modeled as a Poisson distribution as the number of supernovae which occur in a fixed period of time do so independently of each other and form a discrete distribution. The probability of detecting k supernovae is given by

$$P(k;\lambda) = \frac{\lambda^k\ e^{-\lambda}}{k!} \tag{6.2}$$

where λ is the mean number of detections. Zero supernovae were detected hence

$$P(k;\lambda) = \frac{\lambda^0\ e^{-\lambda}}{0!} = e^{-\lambda} \tag{6.3}$$

At the 95% confidence level the mean number of events which would give a zero probability of detecting a supernovae is

$$0.05 = e^{-\lambda} \tag{6.4}$$

$$\lambda = 3.00 \tag{6.5}$$

Therefore at the 95% confidence level 3 detections could have been expected. Using $n = 3$ in Equation 6.1 a rate of 23 deg^{-2} yr^{-1} at the 95% confidence level is found, ignoring cosmic variance uncertainties.

We also calculated the rate in terms of comoving volume of the field. The dark field has a useable area of 214 arcmin^2 or in terms of solid angle 1.81×10^{-5} steradians. Figure 6.8 shows the volume of the dark field across a range of redshifts. Taking into account the 2.2 year baseline the detection rate was converted into units of Gpc^{-3} per year. Figure 6.9 shows the volumetric rate as a function of redshift. The rate was 480 Gpc^{-3} yr^{-1} at $z \leq 5$.

The limits determined here are approximate, as the precise survey area relevant to each individual source is a complicated function of the varying field and depth by epoch. It also doesn't take into account cosmic variance, particularly in relation to low redshift regions of the Universe where metal free Pop III stars would be able to exist. Furthermore, these limits only apply to moderate-duration events, with light curves lasting less than

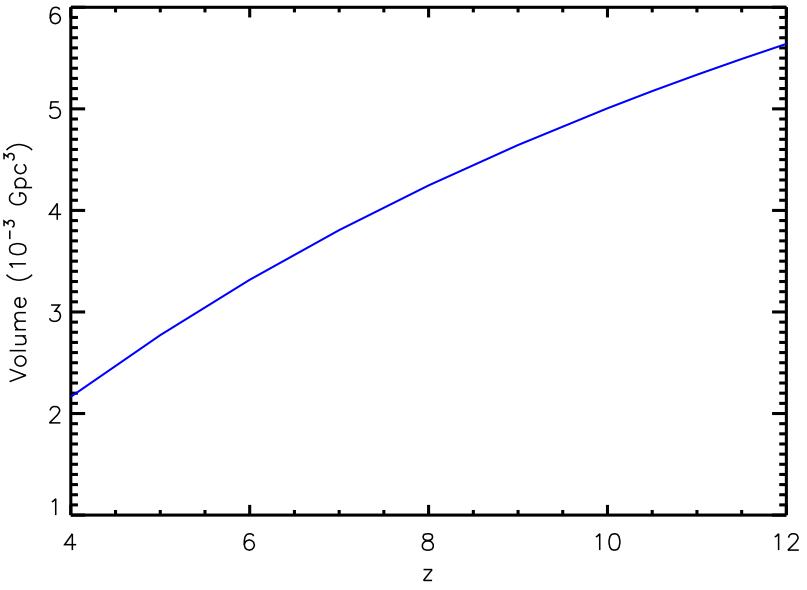


Figure 6.8: The blue solid curve shows the volume of the dark field in 10^{-3} Gpc^3 as a function of redshift.

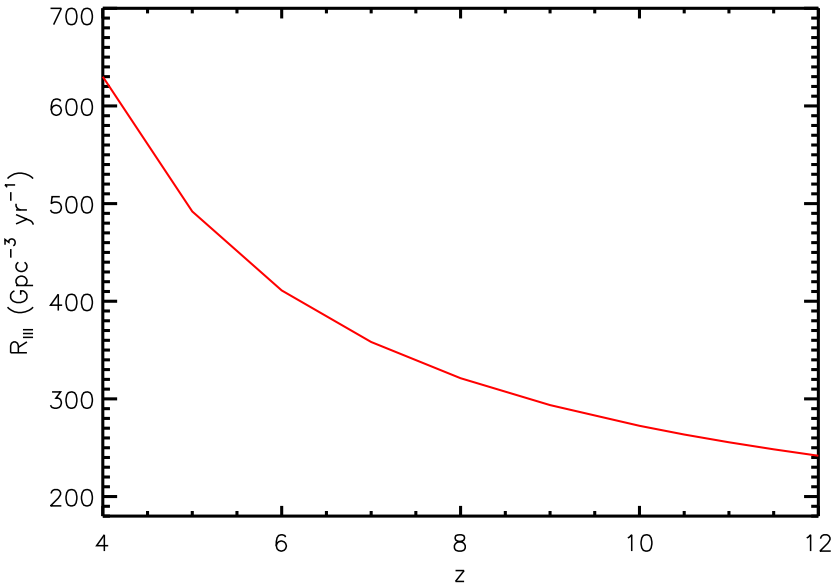


Figure 6.9: The red solid curve shows the 95% confidence level upper limit of the population III supernovae detection rate, R_{III} per Gpc^3 per year as a function of redshift.

$\sim 400/(1+z)$ days, by construction of the search. Many of the predicted light curves given e.g. by Scannapieco(2005) have plateau phases up to 1 yr in the rest frame, which would be too long in the observed frame to be detected by the present survey, but which may be detectable in future incarnations of the *Spitzer* dark field surveys.

6.6.1 Comparison with Predicted Rates

The Wise & Abel(2005), Heger et al.(2002) and Cen(2003) predicted differential rates of $dN/dz \sim 0.34\,\mathrm{deg}^{-2}\,\mathrm{yr}^{-1}$ at $z = 10$ and $\sim 0.2\,\mathrm{deg}^{-2}\,\mathrm{yr}^{-1}$ at $z = 20$ and $\sim 11\,\mathrm{deg}^{-2}\,\mathrm{yr}^{-1}$ at $z > 13$ respectively, are not ruled out, even if they appear at lower redshifts. The rate given by Mackey et al.(2003) is approximately a factor of 2 larger then our rate, however given the uncertainties in our upper limit and the different redshift ranges probed we cannot rule out their model. Given that our search is most effectively probing $z \sim 3-5$, the differential rate of Weinmann & Lilly(2005) of 25 $\mathrm{deg}^{-2}\,\mathrm{yr}^{-1}$ at $z = 5$ is broadly comparable to our limit.

It should be noted, the very luminous PISNe that our search is sensitive to may only be a small fraction of all high-redshift supernova events. There is a distinction currently being made between Pop III.1 and III.2 stars, where the former class are of fully primordial abundance, and form in dark matter mini-halos, resulting in stellar masses of above $\sim 100\,\mathrm{M}_\odot$ which eventually produce PISNe (Johnson & Bromm(2006), McKee & Tan(2008)). These are distinct from the Pop III.2 stars, which are expected to form through atomic cooling processes, producing only $\sim 10\,\mathrm{M}_\odot$ progenitors Greif & Bromm(2006). The PISNe Pop III.1 progenitors may consist of only some 10% of Pop III SNe Greif & Bromm(2006). Furthermore, the "pristine" Pop III.1 progenitors can suffer dramatic negative feedback (e.g. McKee & Tan(2008)), which may additionally limit their relative numbers. These were all considerations not yet taken in the predictions by Mackey et al.(2003), which led to the high expectation rates.

6.7 Summary

In this chapter a search for Pop III SNe was conducted using a unique data set - the IRAC dark field. The dark field has repeat IRAC imaging at 2–3 weeks intervals. Each point in the field typically has more than 10 hours of total integration time, with a maximum of $\sim$ 80 hours in some areas. Systematic searches were conducted for three different

ranges of possible plateau durations. Additional criteria were applied which ensured only reliable data was used. The criteria were broad to ensure no potential candidates were excluded and to take into consideration poorly constrained factors such as the luminosity and redshift of the SNe, the rate of decay of the light curve and the variations in the depth of the field at each point and time. No Pop III SNe were detected and an upper limit was placed on the detection rate of 23 deg^{-2} yr^{-1} at the 95% confidence level. The limit only applies to moderate-duration events, with plateau phases lasting less than $\sim 400/(1+z)$ days. The rate was only approximate as the precise survey area relevant to each individual source was a complicated function of the varying field and depth by epoch.

Chapter 7

Conclusion

In this thesis we investigated the clustering of dark matter haloes and synthetic galaxies from the Millennium Simulation. The clustering of dark matter haloes of the same mass and the evolution of the clustering of haloes was measured over $0 \leq z \leq 1.5$. The clustering of elliptical, spiral and starburst galaxies detected by SWIRE was investigated as a function of stellar mass and redshift. The clustering of spiral galaxies as a function of SFR and redshift was also measured. Finally, a search for population III stars was conducted in the Spitzer dark field.

7.1 Clustering in the Millennium Simulation

The clustering of haloes of the same mass at different redshifts increased for high mass haloes ($\log(M_{hz}/\mathrm{M}_{\odot}) \geq 12.5$) and decreased slightly for low mass haloes ($\log(M_{hz}/\mathrm{M}_{\odot}) \leq 12.0$). We attributed this trend to rarer haloes being more clustered as they tend to be found close to one another.

The clustering of synthetic galaxies from the semi-analytic model of De Lucia & Blaizot(2007) was also measured on large scales. Apart from at $z = 0.09$, more massive galaxies were found to be more clustered. The clustering strength of $\log(M_{\star}/\mathrm{M}_{\odot}) = 10.2$ galaxies decreased over $0.1 < z < 0.5$, whereas $\log(M_{\star}/\mathrm{M}_{\odot}) = 10.8$ galaxies remained constant and $\log(M_{\star}/\mathrm{M}_{\odot}) = 11.2$ galaxies increased. These clustering variations with redshift broadly matched the clustering of the dark matter haloes implying the biasing of such galaxies was constant with redshift. The synthetic galaxies seem too clustered when compared with observational values from the data which we conclude is because the

simulations are too reliant on sustained star formation, meaning massive galaxies are too rare at high redshift and hence reside in the densest environments.

7.2 The Clustering of SWIRE Galaxies

The complicated selection effects inherent to the photometric redshift catalogue from the combination of the SWIRE and INTWFS surveys were thoroughly investigated. The significantly improved understanding of the selection effects will be of great benefit to future work which utilizes this catalogue. Masks were made to remove areas of the fields with low infrared and optical coverage as well as foreground stars. These masks have also been used for other studies, see Oliver et. al. 2009 (submitted). Flux cuts were defined to be used in unison with the masks to minimise variations in completeness to provide a homogeneous data set. n_{band} and χ^2 constraints were defined to ensure only good photometric redshifts were used. A new method was devised to model the artificial variation resulting from the application of these constraints which enabled the variation to be taken into account when the angular correlation function was measured. This enabled us to make maximal use of the catalogues. Stellar mass and SFR incompleteness resulting from the flux and magnitude cuts was investigated and data affected by such incompleteness was not used. The cuts and the masks provided a reliable homogeneous data set with a sufficiently large number of galaxies which had good photometric redshifts.

The photometric redshift catalogue was used to measure the clustering of elliptical, spiral and starburst galaxies as a function of stellar mass out to $z \sim 1.5$. The clustering dependence on star formation rate was also measured for spiral galaxies out to $z \sim 1.5$. Two of the SWIRE fields were used; ELAIS-N1 and ELAIS-N2 covering a total area of 10.34 deg^2. The angular correlation function was measured for each subsample and the comoving spatial clustering strength, r_0, was calculated by inverting through Limber's equation.

We conclude that environmental effects play a major role in the evolution of galaxies. We showed that high redshift ($z = 1.4$) galaxies selected on SFR were more clustered than spiral galaxies of the same stellar mass. Therefore we infer galaxies are star forming because of their environment at these redshifts and that stellar mass does not play an important role. Hence, we conclude environmental effects induce gas instabilities which trigger star formation as opposed to the gas mass being triggered at a point in the galaxies

life almost independently of the environment. We found downsizing was occurring by showing the clustering of SFR selected galaxies decreased toward $z = 0$, a trend which we found was mirrored in decreasing halo mass. The evolution of the halo clustering indicated high z SFR galaxies will evolve into ellipticals by $z \sim 0.5$. This shows it is the galaxies in the densest environments which have their star formation quenched leaving only those star forming galaxies in less dense regions by the present day, hence producing the observed downsizing. This is also supported by the constant relative bias with redshift (at least at $0.6 \leq z \leq 1.0$) which implied that it is the most clustered spirals, from the densest regions, which are turning into ellipticals to preserve the constant relative bias seen here. Therefore, environmentally induced effects are also responsible for turning off star formation, such as galaxy-galaxy interactions which possibly trigger AGN activity (Hopkins et al.(2007)), or processes such as harassment (Moore et al.(1999)) or ram pressure stripping (Gunn & Gott(1972)) which could be responsible for disrupting gas cooling.

These findings coherently show how important galaxy environment is to the evolution of galaxies. It appears to play an important role in not only turning star formation on at high redshift but also quenching it at later times. It is beyond the scope of this work to probe in detail which environmentally induced effects trigger and cessate star formation. This would require a study tailored for such purposes.

The SWIRE photometric redshift catalogue of Rowan-Robinson et al.(2008) has proved to be an excellent data set for measuring the clustering of SWIRE galaxies over $0 \leq z \leq 1.5$. Photometric redshifts have reached a level of maturity through improved template fitting techniques utilizing NIR wavebands, multiple passes through the data and a wider range of templates. Photometric catalogues are now an invaluable resource for high redshift extragalactic studies thanks to coverage spanning multiple fields and large volumes providing large samples of galaxies not attainable through spectroscopic surveys. Now that the complicated selection effects of the SWIRE photometric redshift catalogue are understood enables this dataset to be confidently used not just for more clustering studies but for other future analyses aswell.

7.2.1 Future Clustering Analyses

There is still a lot of clustering work which could be performed with this catalogue such as looking at the clustering of AGN's or clustering as a function of specific star formation rate. One particular area which requires more research is the variation of clustering with

SFR. This could be probed further by incorporating some of the other large SWIRE fields such as the XMM field, if the problems with the photometric redshifts in this field could be improved.

7.3 A Search for Population III Supernovae

A search for Pop III SNe was conducted using a unique data set - the IRAC dark field. The unprecedented depth and multi-epochal nature of the data made it ideal for a first foray to try to detect transient objects which may be luminous pair instability supernovae from the first stars. A database was created combining the epochal data in the field. A bespoke search methodology was designed which took into consideration poorly constrained factors such as the luminosity and redshift of the SNe (from models in the literature) and the variations in the depth of the field at each point and time. Systematic searches were conducted for three different ranges of possible light curve durations including criteria which ensured only reliable data was used. Broad criteria were used to ensure no potential candidates were excluded. HST and IRAC imaging were utilized to identify potential candidates.

Despite the through search method used, no Pop III SNe were detected and an upper limit was placed on the detection rate of 23 deg^{-2} yr^{-1} at the 95% confidence level. This rate was only approximate as the precise survey area relevant to each individual source is a complicated function of the varying field and depth by epoch. This limit only applies to moderate-duration events, with plateau phases lasting less than $\sim 400/(1+z)$ days. Also the search was not sensitive to light curves with long plateau phases of up to 1 yr in the rest frame (Scannapieco(2005)) which would have appeared as continuous sources.

7.3.1 Future Pop III Searches

To do a comprehensive search for Pop III SNe using *Spitzer* (or a future platform with similar capabilities, e.g. JWST) would require a survey area of 1 deg^2 with exposures of 5000 seconds (50×100 s exposures) per point on the sky. This would reach $m_{\mathrm{AB}} \sim 26$ at the 3-5 σ level, depending on extinction. This would remove the problems associated with the survey area relevant to each source. Each epoch would require approximately 200 hours of observation to cover the survey area with a 5 arcminute field of view. The greater depth complicates the IRAC data reduction as confusion is a significant, but tractable

problem. Trading depth for greater area may not be optimal for detecting the *highest* redshift Pop III SNe, which are expected to be quite faint, $m_{\rm AB} \gtrsim 26$. This requirement is increasingly relaxed for lower redshift events, which may possibly exist at redshifts as low as $z \sim 2$. This ideal survey would need to be imaged every two months for a period of several years in order to plausibly sample much or all of the longer duration light curves. The total program would therefore be of the order of 2000-3000 hours. Such a search would be powerful for identifying plausible candidates or at least enabling firmer limits to be set on the production rate of PISNe.

The current dark field dataset continues to expand as *Spitzer* continues to operate, acquiring new IRAC observations every 2–3 weeks. Including observations made during the *Spitzer* Warm Mission, the full dataset will span at least seven years, and maybe as many as ten full years, opening the door for future searches which can probe longer baselines.

Bibliography

Abel, T., Bryan, G. L., & Norman, M. L. 2000, ApJ, 540, 39

Abel, T., Bryan, G. L., & Norman, M. L. 2002, Science, 295, 93

Alimi, J.-M., Blanchard, A., & Schaeffer, R. 1990, ApJL, 349, L5

Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1985, ApJL, 290, L25

Angulo, R. E., Baugh, C. M., & Lacey, C. G. 2008, MNRAS, 387, 921

Babbedge, T. S. R., et al. 2004, MNRAS, 353, 654

Babbedge, T. S. R., et al. 2006, MNRAS, 370, 1159

Baugh, C. M. 1996, MNRAS, 280, 267

Baugh, C. M. 2006, Reports on Progress in Physics, 69, 3101

Bell E. F., 2003, ApJ, 586, 794

Bell, E. F., et al. 2004, ApJ, 608, 752

Bennett, C. L., et al. 2003, ApJ, 583, 1

Berta, S., Fritz, J., Franceschini, A., Bressan, A., & Lonsdale, C. 2004, Astron. Astrophys., 418, 913

Berta, S., et al. 2006, Astron. Astrophys., 451, 881

Berta, S., et al. 2008, Astron. Astrophys., 488, 533

Bertin, E., & Arnouts, S. 1996, Astron. Astrophys., 117, 393

Bett, P., Eke, V., Frenk, C. S., Jenkins, A., Helly, J., & Navarro, J. 2007, MNRAS, 376, 215

Blake, C., Collister, A., Bridle, S., & Lahav, O. 2007, MNRAS, 374, 1527

Blanton, M. R., et al. 2003, ApJ, 594, 186

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2006, MNRAS, 370, 645

Bromm, V., Coppi, P. S., & Larson, R. B. 1999, ApJL, 527, L5

Bromm, V., Kudritzki, R. P., & Loeb, A. 2001, ApJ, 552, 464

Bromm, V., Coppi, P. S., & Larson, R. B. 2002, Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology: Proceedings of the MPA/ESO/MPE/USM Joint Astronomy Conference Held in Garching, Germany, 6-10 August 2001, ESO ASTROPHYSICS SYMPOSIA. ISBN 3-540-43769-X. Edited by M. Gilfanov, R. Sunyaev, and E. Churazov. Springer-Verlag, 2002, p. 316, 316

Bromm, V., & Loeb, A. 2004, New Astronomy, 9, 353

Brown, M. J. I., et al. 2008, ApJ, 682, 937

Bruzual A., G., & Charlot, S. 1993, ApJ, 405, 538

Budavári, T., et al. 2003, ApJ, 595, 59

Calzetti, D., & Kinney, A. L. 1992, ApJL, 399, L39

Carlberg, R. G., Cowie, L. L., Songaila, A., & Hu, E. M. 1997, ApJ, 484, 538

Carlberg, R. G., Yee, H. K. C., Morris, S. L., Lin, H., Hall, P. B., Patton, D., Sawicki, M., & Shepherd, C. W. 2000, ApJ, 542, 57

Cen, R. 2003, ApJ, 591, 12

Cimatti, A., et al. 2004, Nature, 430, 184

Coil, A. L., Newman, J. A., Kaiser, N., Davis, M., Ma, C.-P., Kocevski, D. D., & Koo, D. C. 2004, ApJ, 617, 765

Coil, A. L., Newman, J. A., Cooper, M. C., Davis, M., Faber, S. M., Koo, D. C., & Willmer, C. N. A. 2006, ApJ, 644, 671

Coil, A. L., et al. 2008, ApJ, 672, 153

Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., & Zepf, S. E. 1994, MNRAS, 271, 781

Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168

Cole, S., et al. 2005, MNRAS, 362, 505

Colless, M., et al. 2001, MNRAS, 328, 1039

Cooper, M. C., et al. 2006, MNRAS, 370, 198

Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839

Croton, D. J., et al. 2006, MNRAS, 365, 11

Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303

Cucciati, O., et al. 2006, Astron. Astrophys., 458, 39

Cyburt, R. H., Fields, B. D., & Olive, K. A. 2003, Physics Letters B, 567, 227

Daddi, E., et al. 2003, ApJ, 588, 50

Davis, M., & Geller, M. J. 1976, ApJ, 208, 13

de la Torre, S., et al. 2007, Astron. Astrophys., 475, 443

De Lucia, G., Kauffmann, G., & White, S. D. M. 2004, MNRAS, 349, 1101

De Lucia, G., Springel, V., White, S. D. M., Croton, D., & Kauffmann, G. 2006, MNRAS, 366, 499

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2

Desert, F.-X., Boulanger, F., & Puget, J. L. 1990, Astron. Astrophys., 237, 215

van Dokkum, P. G., et al. 2006, ApJL, 638, L59

van Dokkum, P., & Quadri, R. 2007, arXiv:0704.1760

Draine, B. T. 2003, ARAA, 41, 241

Dressler, A. 1980, ApJ, 236, 351

Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, MNRAS, 235, 715

Efstathiou, G., Kaiser, N., Saunders, W., Lawrence, A., Rowan-Robinson, M., Ellis, R. S., & Frenk, C. S. 1990, MNRAS, 247, 10P

Efstathiou, G., Bernstein, G., Tyson, J. A., Katz, N., & Guhathakurta, P. 1991, ApJL, 380, L47

Elbaz, D. 2005, Space Science Reviews, 119, 93

Espino-Briones, N., Plionis, M., & Ragone-Figueroa, C. 2007, ApJL, 666, L5

Farrah, D., et al. 2006, ApJL, 643, L139

Fazio, G. G., et al. 2004, ApJS, 154, 10

Firth A. E., et al., 2002, MNRAS, 332, 617

Fontana, A., et al. 2004, Astron. Astrophys., 424, 23

Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009, MNRAS, 397, 1776

Fryer, C. L., Woosley, S. E., & Heger, A. 2001, ApJ, 550, 372

Furlanetto, S. R., & Loeb, A. 2003, ApJ, 588, 18

Gabasch, A., et al. 2004, Astron. Astrophys., 421, 41

Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66

Gao, D., Zhang, Y., & Zhao, Y. 2008, Astronomical Data Analysis Software and Systems XVII, 394, 525

Genzel, R., et al. 1998, ApJ, 498, 579

Georgantopoulos, I., & Shanks, T. 1991, The Space Distribution of Quasars, 21, 300

Giavalisco, M., & Dickinson, M. 2001, ApJ, 550, 177

Gilli, R., et al. 2007, Astron. Astrophys., 475, 83

Giovanelli, R., Haynes, M. P., & Chincarini, G. L. 1986, ApJ, 300, 77

Glazebrook, K., et al. 2004, Nature, 430, 181

Gnedin, N. Y., & Ostriker, J. P. 1997, ApJ, 486, 581

Grazian, A., et al. 2006, Astron. Astrophys., 453, 507

Gray, A., et al. 2001, N-body Problems in Statistical Learning, MIT Press

Greif, T., & Bromm, V. 2006, MNRAS, 373, 128

Greif, T., Johnson, J. L., Bromm, V., Klessen, R. S. 2007, ApJ, 670, 1

Gunn, J. E., & Gott, J. R. I. 1972, ApJ, 176, 1

Guzzo L., Strauss M. A., Fisher K. B., Giovanelli R., Haynes M. P., 1997, ApJ, 489, 37

Harker, G., Cole, S., Helly, J., Frenk, C., & Jenkins, A. 2006, MNRAS, 367, 1039

Heger, A., & Woosley, S. E. 2002, ApJ, 567, 532

Heger, A., Woosley, S., Baraffe, I., & Abel, T. 2002, Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology: Proceedings of the MPA/ESO/MPE/USM Joint Astronomy Conference Held in Garching, Germany, 6-10 August 2001, ESO ASTROPHYSICS SYMPOSIA. ISBN 3-540-43769-X. Edited by M. Gilfanov, R. Sunyaev, and E. Churazov. Springer-Verlag, 2002, p. 369, 369

Heinis, S., et al. 2007, ApJS, 173, 503

Hermit, S., Santiago, B. X., Lahav, O., Strauss, M. A., Davis, M., Dressler, A., & Huchra, J. P. 1996, MNRAS, 283, 709

Hinshaw, G., et al. 2009, ApJS, 180, 225

Hogg, D. W., Cohen, J. G., & Blandford, R. 2000, ApJ, 545, 32

Hopkins, P. F., Bundy, K., Hernquist, L., & Ellis, R. S. 2007, ApJ, 659, 976

Houck, J. R., et al. 2004, 5487, 62

Ichikawa, T., et al. 2007, PASJ, 59, 1081

Ilbert, O., et al. 2006, Astron. Astrophys., 457, 841

Iovino, A., Giovanelli, R., Haynes, M., Chincarini, G., & Guzzo, L. 1993, MNRAS, 265, 21

Jenkins, A., et al. 1998, ApJ, 499, 20

Jing, Y. P., Suto, Y., & Mo, H. J. 2007, ApJ, 657, 664

Johnson, J. L., & Bromm, V. 2006, MNRAS, 366, 247

Kaiser, N. 1984, ApJL, 284, L9

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201

Kauffmann, G., Nusser, A., & Steinmetz, M. 1997, MNRAS, 286, 795

Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188

Kessler, M. F., Metcalfe, L., & Salama, A. 1992, Space Science Reviews, 61, 45

Krick, J. E., Surace, J. A., Thompson, D., Ashby, M. L. N., Hora, J. L., Gorjian, V., & Yan, L. 2008, ApJ, 686, 918

Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64

Le Fevre, O., Hudon, D., Lilly, S. J., Crampton, D., Hammer, F., & Tresse, L. 1996, ApJ, 461, 534

Le Fèvre, O., et al. 2004, Astron. Astrophys., 417, 839

Le Fèvre, O., et al. 2005, Astron. Astrophys., 439, 877

Leger, A., & Puget, J. L. 1984, Astron. Astrophys., 137, L5

Lewis, J. R., Bunclark, P. S., Irwin, M. J., McMahon, R. G., & Walton, N. A. 2000, Astronomical Data Analysis Software and Systems IX, 216, 415

Li, A., & Draine, B. T. 2001, ApJL, 550, L213

Li, A., & Draine, B. T. 2001, ApJ, 554, 778

Li, C., Kauffmann, G., Jing, Y. P., White, S. D. M., Börner, G., & Cheng, F. Z. 2006, MNRAS, 368, 21

Lilly, S. J., Tresse, L., Hammer, F., Crampton, D., & Le Fevre, O. 1995, ApJ, 455, 108

Lonsdale, C. J., et al. 2003, PASP, 115, 897

Loveday, J., Efstathiou, G., Peterson, B. A., & Maddox, S. J. 1992, ApJL, 400, L43

Loveday, J., Maddox, S. J., Efstathiou, G., & Peterson, B. A. 1995, ApJ, 442, 457

Lugaro, M., Herwig, F., Lattanzio, J. C., Gallino, R., & Straniero, O. 2003, ApJ, 586, 1305

Mackey, J., Bromm, V., & Hernquist, L. 2003, ApJ, 586, 1

Madgwick, D. S., et al. 2002, MNRAS, 333, 133

Madgwick, D. S., et al. 2003, MNRAS, 344, 847

Maddox, S. J., Efstathiou, G., & Sutherland, W. J. 1996, MNRAS, 283, 1227

Magliocchetti, M., et al. 2008, MNRAS, 383, 1131

McCracken, H. J., et al. 2007, ApJS, 172, 314

McCracken, H. J., Ilbert, O., Mellier, Y., Bertin, E., Guzzo, L., Arnouts, S., Le Fèvre, O., & Zamorani, G. 2008, Astron. Astrophys., 479, 321

McKey, C. F., & Tan, J. C. 2008, ApJ, 681, 771

McMahon, R. G., Walton, N. A., Irwin, M. J., Lewis, J. R., Bunclark, P. S., & Jones, D. H. 2001, New Astronomy Review, 45, 97

Meneux B., et al., 2006, A&A, 452, 387

Meneux, B., 2007, A&A, 11, 20070629

Meneux, B., et al. 2008, Astron. Astrophys., 478, 299

Meneux, B., et al. 2009, arXiv:0906.18

Miralda-Escude, J., & Rees, M. J. 1997, ApJL, 478, L57

,Moore, A. W., Lee, M. S., Volume 8 of Journal of Artificial Intelligence Research

Moore, B., Lake, G., Quinn, T., & Stadel, J. 1999, MNRAS, 304, 465

Moster, B. P., Somerville, R. S., Maulbetsch, C., van den Bosch, F. C., Maccio', A. V., Naab, T., & Oser, L. 2009, arXiv:0903.4682

Neugebauer, G., et al. 1984, ApJL, 278, L1

Nilson, P. 1995, VizieR Online Data Catalog, 7026, 0

Norberg, P., et al. 2001, MNRAS, 328, 64

Norberg, P., et al. 2002, MNRAS, 332, 827

Oliver, S., Serjeant, S., Efstathiou, A., Crockett, H., Gruppioni, C., La Franca, F., & The Elais Consortium 2000, ISO Survey of a Dusty Universe, 548, 28

Oliver, S., et al. 2000, MNRAS, 316, 749

Omukai, K., & Palla, F. 2003, ApJ, 589, 677

O'Shea, B. W., McKee, C. F., Heger, A., & Abel, T. 2008, ArXiv e-prints, 801, arXiv:0801.2124

Overzier, R. A., Röttgering, H. J. A., Rengelink, R. B., Wilman, R. J., 2003, Astron. Astrophys., 405, 53

Padmanabhan, N., et al. 2007, MNRAS, 378, 852

Panuzzo, P., Silva, L., Granato, G. L., Bressan, A., & Vega, O. 2005, The Spectral Energy Distributions of Gas-Rich Galaxies: Confronting Models with Data, 761, 187

Peacock, J. A., 1999, Cosmological Physics, (Cambridge, CUP)

Peacock, J. A., & Smith, R. E. 2000, MNRAS, 318, 1144

Peacock, J. A., et al. 2001, Nature, 410, 169

Peebles, P. J. E. 1980, Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.,

Percival, W. J., et al. 2001, MNRAS, 327, 1297

Percival, W. J., et al. 2007, ApJ, 657, 645

Perlmutter, S., et al. 1999, ApJ, 517, 565

Phleps, S., Peacock, J. A., Meisenheimer, K., & Wolf, C. 2006, Astron. Astrophys., 457, 145

Phillipps, S., Fong, R., Ellis, R. S., Fall, S. M., MacGillivray, H. T., 1978, MNRAS, 182, 673

Poggianti, B. M., Bressan, A., & Franceschini, A. 2001, ApJ, 550, 195

Pollo, A., et al. 2006, Astron. Astrophys., 451, 409

Pope, A. C., et al. 2004, ApJ, 607, 655

Quadri, R., et al. 2007, ApJ, 654, 138

Quadri, R. F., Williams, R. J., Lee, K.-S., Franx, M., van Dokkum, P., & Brammer, G. B. 2008, ApJL, 685, L1

Riess, A. G., et al. 1998, AJ, 116, 1009

Riess, A. G., et al. 2004, ApJ, 607, 665

Rieke, G. H., & MIPS Team 2004, Bulletin of the American Astronomical Society, 36, 700

Roche, N., Eales, S. A., Hippelein, H., Willott, C. J. 1999, MNRAS, 306, 538

Rowan-Robinson, M., et al. 1997, MNRAS, 289, 490

Rowan-Robinson, M. 2001, ApJ, 549, 745

Rowan-Robinson M., 2003, MNRAS, 344, 13

Rowan-Robinson, M., 2005, ApJ, 129, 1183-1197

Rowan-Robinson, M., et al. 2008, MNRAS, 386, 697

Saunders, W., Rowan-Robinson, M., & Lawrence, A. 1992, MNRAS, 258, 134

Scannapieco, E., Madau, P., Woosley, S., Heger, A., & Ferrara, A. 2005, ApJ, 633, 1031

Schaerer, D. 2002, Astron. Astrophys., 382, 28

Schaerer, D. 2003, Astron. Astrophys., 397, 527

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Seldner, M., Siebers, B., Groth, E. J., & Peebles, P. J. E. 1977, AJ, 82, 249

Shankar, F., Lapi, A., Salucci, P., De Zotti, G., & Danese, L. 2006, ApJ, 643, 14

Shepherd C. W., Carlberg R. G., Yee H. K. C., Morris S. L., Lin H., Sawicki M., Hall P. B., Patton D. R., 2001, ApJ, 560, 72

Sicotte, H., & Peebles, P. J. E. 1995, Dark Matter, 336, 390

Skrutskie, M. F., et al. 2006, AJ, 131, 1163

Smith, N., et al. 2007, ApJ, 666, 1116

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726

Springel, V., et al. 2005, Nature, 435, 629

Springel, V. 2005, MNRAS, 364, 1105

Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137

Strateva, I., et al. 2001, AJ, 122, 1861

Strauss, M. A., Davis, M., Yahil, A., & Huchra, J. P. 1992, ApJ, 385, 421

Swanson, M. E. C., Tegmark, M., Blanton, M., & Zehavi, I. 2008, MNRAS, 385, 1635

Tegmark, M., et al. 2004, ApJ, 606, 702

Tornatore, L., Ferrara, A., & Schneider, R. 2007, MNRAS, 382, 945

Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 119, 1645

Trumpler, R. J. 1930, PASP, 42, 214

Tumlinson, J., & Shull, J. M. 2000, ApJL, 528, L65

Umeda, H., & Nomoto, K. 2002, ApJ, 565, 385

Waddington, I., et al. 2007, MNRAS, 381, 1437

Warren, S. J., Iovino, A., Hewett, P. C., & Shaver, P. A. 1993, Observational Cosmology, 51, 163

Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71

Weingartner, J. C., & Draine, B. T. 2001, ApJ, 548, 296

Weinberg, D. H., Davé, R., Katz, N., & Hernquist, L. 2004, ApJ, 601, 1

Weinmann, S. M., & Lilly, S. J. 2005, ApJ, 624, 526

Werner, M. W., et al. 2004, ApJS, 154, 1

Wetzel, A. R., Cohn, J. D., White, M., Holz, D. E., & Warren, M. S. 2007, ApJ, 656, 139

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52

Willmer, C. N. A., da Costa, L. N., & Pellegrini, P. S. 1998, AJ, 115, 869

Wise, J. H., & Abel, T. 2003, The Emergence of Cosmic Structure, 666, 97

Wise, J. H., & Abel, T. 2005, ApJ, 629, 615

Woosley, S. E., Blinnikov, S., & Heger, A. 2007, Nature, 450, 7168

Xu, C. K., Lonsdale, C. J., Shupe, D. L., Franceschini, A., Martin, C., & Schiminovich, D. 2003, ApJ, 587, 9

Yang, X. H., Mo, H. J., Kauffmann, G., & Chu, Y. Q. 2003, MNRAS, 339, 387

York, D. G., et al. 2000, AJ, 120, 1579

Yoshii, Y., & Takahara, F. 1988, ApJ, 326, 1

Yoshida, N., Omukai, K., Hernquist, L., & Abel, T. 2006, ApJ, 652, 6

Zehavi, I., et al. 2002, ApJ, 571, 172

Zehavi, I., et al. 2004, ApJ, 608, 16

Zhu, G., Zheng, Z., Lin, W. P., Jing, Y. P., Kang, X., & Gao, L. 2006, ApJL, 639, L5